

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 40.6668 Seconds  
(without alignments) 2761.097 Million cell updates/sec

Title: US-09-647-544-2  
Perfect score: 6106  
Sequence: 1·MELEPFTVTHLEPLVFTGLC.....GFFAHKKIPPEEKREEKLEQ 1167

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_79:\*

```
1: pir1:*
```

```
2: pir2:*
3: pir3:*
```

```
3: p1r3:*
4: p1r4:*
```

4: p1r4::

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1894.5	31.0	90	1180	2	A35954	integrin alpha-1 c
2	1887	30.9	89	1181	2	A45226	integrin alpha-1 c
3	1756	28.8	88	1181	2	A33998	integrin alpha-2 c
4	1754	28.7	88	1170	2	I45914	integrin alpha-2 s
5	1744	28.6	88	1178	2	S44142	VLA-2 protein homo
6	1105	18.1	1170	2	S03308	cell surface glyco	
7	1097	18.0	1163	2	I56126	lymphocyte function	
8	1085.5	17.8	1183	1	RWHUB	cell surface glyco	
9	1054	17.3	1163	1	RWHUC	cell surface glyco	
10	1022.5	16.7	1153	2	S00551	leukocyte surface	
11	932.5	15.3	1179	2	A53213	integrin alpha-E c	
12	738	12.1	1035	2	I58409	integrin alpha-9 c	
13	732.5	12.0	1054	2	JC7294	integrin -	
14	690	11.3	1039	2	A41131	lymphocyte-Peyer's	
15	684	11.2	1038	2	S06046	integrin alpha-4 c	
16	661	10.8	1041	2	T31437	integrin alpha-4 c	
17	655	10.7	272	2	A55348	integrin alpha-1 -	
18	640.5	10.5	1137	2	JC5950	integrin alpha-7 c	
19	602.5	9.9	1051	2	A40021	integrin alpha-3 c	
20	600	9.8	1031	2	A41543	integrin alpha-6 c	
21	598.5	9.8	1073	2	B36429	integrin alpha-6 c	
22	596.5	9.8	1072	2	A38457	integrin alpha-6 c	
23	594.5	9.7	1135	2	I61186	alpha-7 integrin -	
24	584.5	9.6	1044	2	T10050	integrin alpha-v c	
25	581.5	9.5	1053	2	I55534	VLA-3 alpha subuni	
26	576.5	9.4	1044	2	I65116	integrin alpha-8 c	
27	576.5	9.4	1051	2	A35761	cell surface glyco	
28	574	9.4	1048	2	A27421	integrin alpha-5 c	
29	573.5	9.4	1106	2	S38783	integrin alpha-4 c	

## ALIGNMENTS

## RESULT 1

A35854

integrin alpha-1 chain precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 23-Oct-1990 #sequence\_revision 13-Sep-1991 #text\_change 09-Jul-2004  
C;Accession: X25854. C11242

C;Accession: A35854; S11243  
P.Tanative M.J : large T

J. Cell Biol. 111: 709-720, 1990.

A>Title: Molecular cloning of the rat integrin alpha-1-

A;Title: MOLECULAR CROWDING OF THE RAT INCEPHALIN ALPHA 1  
A;Reference number: A35854; MUID:90338125; PMID:23802499

A;Accession: A35854

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1180 <IGN>

**Keywords:** cell adhesion; cytoskeleton; transmembrane protein A; Cross-references: UNIPROT:P18614; GB:X52140; NID:g56493; PIDN:CAA36384.1; PID:g56494

**C;key words:** cell adhesion; cytoskeleton; transmembrane protein

F, 170-343/DOMAIN: von witterband factor type A repeat homology <vnaaz>

Query Match	31.0%;	Score 1894.5;	DB 2;	Length 1180;
Best Local Similarity	36.5%;	Pred. No. 1.6e-132;		
Matches 436;	Conservative 221;	Mismatches 461;	Indels 77;	Gaps 24;

Qy	13	LVLFTGLCSFPNLDHEHPRLFPGPPEABFGSVLQHVGGQGRVMVLGAPWDPGSDRRGD	72
Db	19	LTVILGFCVFNVDVQKNSMFSFGVEBDMFGYTVQYENEEKGWYLGSLVYGQPKARTGD	78
Qy	73	VYRCPVGAHNA PCAKGHLDGYQLGNSSHAV-----NMHLGMSLLETDGGGGMACAPL	127
Db	79	VYRCPVGERAMPCKVLDLP-----VNTSI PNVTIEIKENMTFG-STLVNPNMGGLUACGPL	133
Qy	128	WSRACSSVFSSGGICARVDASFOQGS LAPTAQRCPTVMYDVI VLDGNSNISPSEVOTF	187
Db	134	YAYRCGHLHYTTGICSDVSTFQVNVSFAP-VQECSTQLDIVI VLDGNSNISPSEVATF	192
Qy	188	LRRLVKGKLFIDPQIOVGLVOYGESPVHWSLGBFRTKEEVVRAAKNLSRREGRETAKQ	247
Db	193	LNDLKRMIDGPKQTQVGI VOYGENVTHEFNLNKYSSTEVELVAANKIGROGGLQTM TAL	252
Qy	248	AIWVACTGESQSHGGPEARLLVVVTDGESHDPGEELPAAL KACEAGRVTTRYGI AVLGH	307
Db	253	GIDTARKEAFTEARGARGVKVMVITVDGESHDPNYRLKQVIOCEPENIORFSI AILGH	312
Qy	308	YLRRQRPDSFLREIRTIASDPDERFPFNVTTDEAALTDI VDALGDRITFGLBSHAENESS	367
Db	313	YNRGNLSTFKPVEIKSIASEPTKHPFNVSDELA LVTIVKALGERIFALEATADQSAAS	372
Qy	368	FGLEMSQIGFSTHRLKDGILFGMWGAYDWGGSVLWLEGGHRLFPFRMALEDFPPALQNH	427
Db	373	FEMEMSQTGFSAHYSQDWVWMLGAVGYDWNGTVMQKQANQMVIPHNTTFQTE--PAKXNE	430
Qy	428	--AAYLGYSVSSMLLRGRRLFTLSGAPFRFRHRGKVIAFQLKKDGAVRVVAQSLOQEQTGSY	485

```
Db 431 PLASYLYGTVNSATIPGD-VLYIAGQPRYNHTGQVVIYKM-EDGNINILQTLGEGQISY 488
Qy 486 FGSLCPLDTRDQTTDVLVAAPFLGPONKGTGRVYVYLVGQOQLLTQLGTLP---- 541
Db 489 FGSVLTTIDDKSYTDLVLGAPMYMGTEKEGQKYVYAV-NQTFEYQMSLEPIRQT 547
Qy 542 -----EPPQDARFGFAMGALPDLNQDGFADVAVGAPLEDHGQGHQALYLYHG 586
Db 548 CCSLKDNSCTKENKNEPCGARGFATAA VKDLNVDFNDVWIGAPLEDHAGAVIYHG 607
Qy 587 TQSGVRPHPAQRIAAASMPHALSVFGRSVDRGLDGDGLVDVAVAGQAAILSSRPV 646
Db 608 SGTIREAYAQRIIPSGDGTLPFGQSIHGMELNGDGLTDTVITIGLGAALFWARDVA 667
Qy 647 HLTPSLVTPQASVSVORDCRRRQGEAVCLTAALCFQVTSKTPGRMBHQFMRTASLDE 706
Db 668 VVKVTMNFENKVNQKNCRVEGETVCINATCFHVKLKSKEDSIYEADLQVRYTLD 727
Qy 707 WTAGARAAPDGSQRLSPRLRLSVGNVTCBQLHFHVLDTSDYLRLPVALTVPFALDNTTK 766
Db 728 LRQISRSFFSGTQERKIQR--NITVRESECIHRSFYMLDKHDFQDSVRVTLDF---NLTD 782
Qy 767 P--GPVLNBSPTSQIKLFPSCDGDNECVTDVLQVNMDIRGSKAFVVRGGRKV 824
Db 783 PENGVPVLDALPNSVHEHIPPACDGNKERCISDLTLNVST---TBKSLLIIVKSKDXF 838
Qy 825 LVSTTLNKRKENAYNTSLSIIFSRNLHLASLTP-ORESPIKVECAAPSAHARLCSVGHV 883
Db 839 NVSLTVKXKGDASNTVTVQHSNLIIPSGIEELQKDS-----CESNQNTCRGVYPP 891
Qy 884 FQTGAKYTFLLPEFSCSSLLSQVFGKITASSDSLRNGTLQENTTAQTSAYIQYEPHLF 943
Db 892 LRAGETVTFKIIQFNTSHLSENALIHLSATSDSEEPLESNDNEVNSIPVKYEVGLQF 951
Qy 944 SSESTLRYEYHPVGTLP-----VGPQPEKTLRVQNLGCVVSGLIISALLP-AVAH 996
Db 952 YSSASEHHISVAANETIPEFINSTEDIGNEINVFYTIKRGHPMPPELQLSISPPNLTD 1011
Qy 997 GGNVFLSLQSVITNASCIVQNLTEPPG-----PPVHPELOHTNRLNGSNTOCQVVR 1050
Db 1012 GYPVLYPIGWSSDUNVCRPSLEDPFGINSKGKQMTISKSEVLKRGTIQDCSSTCGVATI 1071
Qy 1051 HLGQAKG-TEVSVGLRLVHNEFFRRAKFKSLTVVSTFELGTBEGSVLQTEASRWSES 1109
Db 1072 TCSLLPSDLSQVNSL--LLWKPTFIRAHFSSNLTLRGLKSENSS-LTLSSNRKREL 1128
Qy 1110 LLEVQOT-RPILISLWILIGSVLGGILLALLVFLWKLVGFAFKKIPBEKREE 1163
Db 1129 AIQISKDGLPGRVPLWILLSAFAGLLLLMLLILALWKIGFF---KRPLKKOMEK 1180

RESULT 2
Integrin alpha-1 chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A45226
R:Briesewitz, R.; Epstein, M.R.; Marcantonio, E.E.
J. Biol. Chem. 268, 2989-2996, 1993
A:Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.
A:Reference number: A45226; MUID:93155124; PMID:8428973
A:Accession: A45226
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Releases: 1-1151 <BRI>
A:Cross-references: UNIPROT:P56199
A:Experimental source: hepatoblastoma cell line HepG2
A>Note: sequence extracted from NCBI backbone (NCBIP:124326)
F:142-317/Domain: von Willebrand factor type A repeat homology <WVAL>

Query Match 30.9%; Score 1887; DB 2; Length 1151;
Best Local Similarity 36.3%; Pred. No. 5.4e-132;
Matches 429; Conservative 226; Mismatches 454; Indels 74; Gaps 22;
```

```
Qy 23 FNLDHHPRLFPGPPEAEFGYSVLQHVGGQRMVLGAPWDGSGDRRGDGVYRCPVGGAH 82
Db 1 FNVDVKNSMTFSGPVEDMGFTVQQYENBEGKWLIGSPLVGQPKNKTGDKVYKCPVGRGE 60
Qy 83 NAPCAKGHLDYQJLGNSSHPAV-----NMHLGMSLLETGDDGDFMACAPLWSRACSSVF 137
Db 61 SLPCVKLDLP-----VNTSI PNVTVEKENMTFG-STLVNPNNGGFLACGLPYAYRCCGHLHY 115
Qy 138 SSGICARVDASFPQOGLAPTAQRCPYMDVIVLDGNSIYPMWSEVQTFRLRLVGKLEI 197
Db 116 TTGICSDVSPTFQVNNSIAP-VQSCSTQLDIVIVLDGNSIYPMWSTVTAFLNDLLKRMDI 174
Qy 198 DPEIQIQLVQYGESPVHWSLGDFTKBEVRAAKNLNRREGRETKTQAIAINVACTEGF 257
Db 175 GPKQTQVGIQYGVENVTHEFNLYKSYSTEVLVAAKIVQRCGRQWTALTGTARKEAF 234
Qy 258 SQSHGRRPBAARLLVVVTGESHGDBELPAALKACBAGRVTYRGIYAVLGHYLRQRDPSS 317
Db 235 TEARGARRGVKVMVITDGHSDHNRLLKKVIOQCEDENIQRPISAILGSYNRLSTEK 294
Qy 318 FLREIETIASDPPERFENVTDAAALTDIVDALGDRIEGLGSHAENESSFGLEMSQJGF 377
Db 295 FVEEIKSIASEPTEKHFENVSDDELAVTIVKTIGERIFALEATADQSAASFEEMSQTGF 354
Qy 378 STHRLKDGILFGMVGAYDMGGSVLMLEGGHRLPPRMALEDEPPALQNHAAVLYGSVSS 437
Db 355 SAHYSQDMVLMGAVGAYDMNGTVVMQKASQIIIPRNTTFNVESTKKEPLASLYGTVNS 414
Qy 438 MLURGRRRLPLSGAPRHRGKVIAPQLKKGAVRVAQSLQGGQISYFGSELCPLDTR 497
Db 415 ATASSGDVLYIAQPRYNHTGVIIYRM-EDGNIKILQTLISGQISYFGSILTTTDDIK 473
Qy 498 DGTDTLLVAAPNPLGPONKETGRVVVYLVGQOQLLTQLGTLP----- 541
Db 474 DSNTDILLVGAPMYMGTEKEBQKVYVYAL-NQTFEYQMSLEPIKQTCSSRQHSNCTT 532
Qy 542 ---EPQDARFGFAMGALPDLNQDGFADVAVGAPLEDHGQALYLYHGTSQGVYRHPAQR 598
Db 533 ENKNEPCGARFGTAIAAVKDLNLDGNDIVIGAPLEDHAGAVIYHSGKTKIRKAYQR 592
Qy 599 IAAASMPHALSVFGRSVDRGLDGDGLVDVAVAGQAAILSSRPVILHTPSLEVTPOA 658
Db 593 IPSGDGTKLFPQGIHGMELNGDGLTDTVITIGLGAALFMSRDVAVVVKVMTFEPNK 652
Qy 659 ISVVQRDCRRRQGEAVCLTAALCFQVTSRTPGRMDHQFYMRFTASLDEWTAGARAFDGS 718
Db 653 VNTQKNCHMEGKETVCINATVCFEVKLSKEDTIYEADLQYRVTLDSLRLKQISRSFSGT 712
Qy 719 GORLSRRRLRLSVGNVTCBQLHFHVLDTSDYLRLPVALTVPFALDNTTKP--GPVLNBSGP 776
Db 713 QERKVQR--NITVRKSECTKHSFYMLDKHDPQDSVRITLDF---NLTDPENGFVLDLSP 767
Qy 777 TSOKLVPSKCGCPDNECVTDVLQVNMDIRGSKAFVVRGGRKVLVSTTLNKRKEN 836
Db 768 NSVHEVTPFAKDCGKCEKICISDLSHV-----ATTEKDLIVRSQNDKFNVSILTQNTKDS 823
Qy 837 AYNTSLSIIFSRNLHLASLTPQRESPIKVECAAPSAHARLCSVGHVPFQAGKVTFLLEF 896
Db 824 AYNTRIYHVSPLNLVFSGI-----EATQKQSC--ESNHNITCKYGVPLRREGVMVTKILF 877
Qy 897 EFGCSLLSQVFGKLTASSDSLRNGTLQENTTAQTSAYIQYEPHLLFSSSTLHRYEVHP 956
Db 878 QFNTSYLMENVTIYLSATSDSEPEPTELSNVNVISIPVKYEVGLQFYSASEVHISIAA 937
Qy 957 YGTLP-----VGGPGEFKTLRVQNLGCVVSGLIISALLPAVAGGN---YFLSLSQV 1007
Db 938 NETVPEINSTEIDIGNEINIFYLIRKSGSPMPELKLISFPNMTSGYPLVPTGLSS- 996
Qy 1008 ITWNASCIQVNLTEP-----PGPPVHPELOHTNRLNGSNTOCQVVRCHLQGLAKGTEV 1061
Db 997 -SENANCRPHIFDPFINSKGKQMTSTDLHKKRGTLDCNCTCKPATITCMLTS-SDISQV 1054
```

QY 1062 SVGLRLVHNEFFRRRAKFKSLTVVSTELCTEGSVLQLTASRWSLSLEVVQT-RPIL 1120  
 Db 1055 NVSL--LLMKPTFKTSYFSSNLTIRGELSENAS-LVLSNNOKRELAIQISKDGLPGR 1111  
 QY 1121 ISWLILGSLVGLGLLLALLVFCILKGLGFFFAHKKIPPEEKREE 1163  
 Db 1112 VPLWVILLSAPAGLLMLLJLALWKIGFF---KRPLKKKWEK 1151

RESULT 3  
 A33998  
 N;Integrin alpha-2 chain precursor - human  
 N;Alternate names: CD49b; platelet glycoprotein GPIa; VLA-2/collagen receptor alpha-2 chain; Species: Homo sapiens (man)  
 C;Date: 30-Mar-1990 #sequence\_revision 18-Sep-1992 #text\_change 09-Jul-2004  
 C;Accession: A33998; B56793; A53117  
 R;Takada, Y.; Hemler, M.E.  
 J. Cell Biol. 109, 397-407, 1989  
 A;Title: The primary structure of the VLA-2/collagen receptor alpha (2)-subunit (platelet)  
 A;Reference number: A33998; PMID:89308879; PMID:2545729  
 A;Accession: A33998  
 A;Molecule type: mRNA  
 A;Residues: 1-1181 <TAK>  
 A;Cross-references: UNIPROT:P17301; GB:X17033; NID:q33906; PIDN:CAA34894.1; PID:q33907  
 A;Note: The authors translated the codon GAT for residue 802 as Gln, GTC for residue 803  
 R;Catimel, B.; Parentier, S.; Leung, L.L.; McGregor, J.L.  
 Biochem. J. 279, 419-425, 1991  
 A;Title: Separation of important new platelet glycoproteins (GPIa, GPIC, GPIc\*, GPIIa and GPIb)  
 A;Reference number: A56793; PMID:92061944; PMID:1953640  
 A;Accession: B56793  
 A;Molecule type: protein  
 A;Residues: 30-43 <CAR>  
 A;Experimental source: platelet  
 R;Zutter, M.M.; Santoro, S.A.; Painter, A.S.; Teung, Y.L.; Gafford, A.  
 J. Biol. Chem. 269, 463-469, 1994  
 A;Title: The human alpha-2 integrin gene promoter. Identification of positive and negative regulatory elements  
 A;Reference number: A53117; PMID:94103255; PMID:8276836  
 A;Accession: A53117  
 A;Molecule type: DNA  
 A;Residues: 1-16, 'V', 18-21 <ZUT>  
 A;Cross-references: GB:L24121; NID:g400342; PIDN:AAA16619.2; PID:g4583535  
 A;Note: authors translated the codon GTA for residue 17 as Leu  
 C;Genetics:  
 A;Gene: GDB:ITGA2; CD49B  
 A;Cross-references: GDB:I28031; OMIM:192974  
 A;Map position: 5q11.1-5q11.2  
 C;Keywords: cell adhesion; glycoprotein; heterodimer; transmembrane protein  
 F;1-29/Domain: signal sequence #status predicted <SIG>  
 F;30-1133/Domain: extracellular #status predicted <EXT>  
 F;172-347/Domain: von Willebrand factor type A repeat homology <VWA2>  
 F;1134-1154/Domain: transmembrane #status predicted <TM>  
 F;1155-1181/Domain: intracellular #status predicted <CYT>  
 F;105,112,343,432,460,475,699,1057,1074,1081/Binding site: carbohydrate (Aen) (covalent)

Query Match 28.8%; Score 1756; DB 2; Length 1181;  
 Best Local Similarity 34.3%; Pred. No. 3.4e-122;  
 Matches 418; Conservative 219; Mismatches 455; Indels 126; Gaps 31;

QY 11 LPLVFLTLG-----CSPFNLDHHPRLFPGRPEAEFGYVLOHVGSGGRWMLVGAPWD 63  
 Db 11 LPLLLVLLSQILNCLLAYNVGLPEAKIFGSPSEQGYAVQVQFIPKGNWLLVGSFWS 70  
 QY 64 GPSGRRGDVYRCPVGGAAHAPCAKGLH-GDYQLGNSHPAVNMHLGMSLLLETDDGGCFM 122  
 Db 71 GFPENRMGDVYKCPV-DLSTATCEKLNLTQSTSPNVTEMKTNMSLGLILTRNMGTCGFL 129  
 QY 123 ACAPLWSRACSSVPSSGICARVDASFPQCSLAPTAQRCPTYMDVIVLDCGNSIYFWS 182  
 Db 130 TCGPLWAAQQCGNQYVITGVCSDISPDFQLSASFSPATQCFSLIDVVVVCDESNIYFWD 189  
 QY 183 EVQTFLLRLVGLKFLTDPDQIQVGLVQYGESPVHEWSLGDFTKERVVRAAKNLSRREGRE 242  
 Db 190 AVKNFLEKFOGLDIGNTKTQVGLIQYANNPRVFNLTNTYKKEIMVATSTQSYGGDL 249

RESULT 4

I45914

Integrin alpha 2 subunit - bovine (fragment)

QY 243 TKTAQAIMVACTGFSQSHGCRPEEARLLVVVTDGESHGDEELPAALKACAGRVTRYGI 302  
 Db 250 TMTFGAIQYARKYAYSAAGRRSATKVMVVVTDGESHGSMKLVAVDQCNDHNDILRPGI 309  
 QY 303 AVLGCHYLRQRDPSSFLUREIRTIASDPDEFNFNVNTEAALTIDVLDGRIFCLESHA 362  
 Db 310 AVLGYNLRNALTNYLKEIKATASIFTERYFFNVSDAALLEKAGTLGEOIFIEGT-V 368  
 QY 363 ENSSFGLEMSQIGFST--HRLKDGILFGVMGAYDMGGSVLW-LEGHRLFPFPPMALEDE 419  
 Db 369 QGDNFQEMSVQVFSADYSYQNDILMLGAVGAFGWSGTIVQKTSHGHLIFP-----KQA 423  
 QY 420 FPAALQ--NHAAYLGVSVSMILRGRRFLSGAPRFRHGKVIAPOLKDGAVRVQASL 477  
 Db 424 FDQILQDRNHSSVLYGSVAA-ISTGESTHFVAGAPRANYTGOIVLYSVNENGNIYIAH 482  
 QY 478 QGQIGSYFGSELCLPLDTRDGTDLVLLVAPMFLGPONKETGRVYVY-----LVGQOQL 532  
 Db 483 RGOIGSYFGVLCSDVDVDDTTDVLVAGAPMTMSDLKKBEGRVYLFTTKKGLGQHOF 542  
 QY 533 LTLQGLTQPEPPQDAREFGFAMGALPDNLNQDFADVAVGAPLEDHGQALYLYHGTQSGVR 592  
 Db 543 --LEG---PSEIENTRFGSAIALSDINMDGENDVIVGSPLENQNSGAVIYNGHQGTIR 597  
 QY 593 PHPAQRIAAA--SMPHALSIFGRSDGRDLDDGDLVDVAVGAQAAAILSSRPVHLTP 650  
 Db 598 TKYSQKILGSDGAFRSHLQYFGRLSDGYGLNGDSITDVSIGAFGVQVQLMSQSIADVAI 657  
 QY 651 SLEVTQPAISVQORDCHRRRQEAVALCFAQVTSRTPGRDWHQFVMRTASTASDEWTAG 710  
 Db 658 EASFTEPKITLVNKNQ-----ILKLCFSAKFR-PTKQNNQVAIVNITLDA----704  
 QY 711 ARAAFDGSQORLSPPRL-----RLSVGNV-----TCEQLHFHVLDTSDYLRPVALTVT 758  
 Db 705 -----DGFSSRVTSRGLFKENNERCLQKNVNVQAQSCPEHIIYIQSPSDVNVSLDLRVD 759  
 QY 759 FALDNTTKPG--PVLNEGSPITSQKLVPSKDCPDNECVTDLVLQVNMIDIRGSRKAPV 816  
 Db 760 ISLEN-----PGTSPALAEYSETAKVFSIPFHKDCGEDGLCISDLVDVR-QIPAAQEQPFI 815  
 QY 817 VRGRRKVLVSTLTLENKENAYNTLSIIFSRNHLASLTPQRESPI--KVEC-AAPSA 872  
 Db 816 VSNQKRLTFSVLTKNKRSAINTGIVDFSENLFPAFS-----LPVDGTEVTCQVAASQ 871  
 QY 873 HARLCSVGHVFOFGAKVTFLEPEFSCSILLSQVFGKLTASSDSLERNGTLOENTAQTS 932  
 Db 872 KSVACDVGYPALKREQQVTTINFDNLQNLQOASLSFOALSSESQENKA--DNLVNLK 929  
 QY 933 AYIQYEPHLLFSSSESTLHRYEVHPYGLP-----VGPGEFFKTLRVQNLGCVVSGLI 986  
 Db 930 IPLLVDAAEIHILTRSTNINFEISDGNVPSIVHSFEDVGPKFIFSLKV-TTGSVPVSMAT 988  
 QY 987 ISALLPAVAGGVNPLSLSOVITNNASCIQNLTEPP-----GPPVHPEELQHTNRL 1038  
 Db 989 VIHIPOYTKKPNMPLMTGTQTDKAGDISCNADINFLKIQTSSSSVFKSENFRHTKEL 1048  
 QY 1039 NGSNTQCVVVRCHLQOLAKGTEVSGLLRLVHNEFFRRAKFKSLTVVSTFELGTEGSVL 1098  
 Db 1049 NCRTASCSNVTCLWKDVHMKGEYFVNVTRIWNGTFASSITFQTVOLTAABAINVYPIY 1108  
 QY 1099 QLTEASRWSLELVQVTRPILISLWI-----LIGSVLGLLLALLLALLVCLWL 1145  
 Db 1109 -----VIEDNTVTIPLMIMKPKAEVPTGVIIGSIAGIALLLALLVAILW 1154  
 QY 1146 KLGFPAHK-----KIPES 1158  
 Db 1155 KLGFKKRYKWKTKNPDE 1172

C;Species: Bos primigenius taurus (cattle)  
 C;Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 09-Jul-2004  
 C;Accession: I45914  
 R;Kamata, T.; Puzon, W.; Takada, Y.  
 J. Biol. Chem. 269, 9659-9663, 1994  
 A;Title: Identification of putative ligand binding sites within the I-domain of integrin alpha5beta1  
 A;Reference number: A54402; MUID:94193647; PMID:7511592  
 A;Accession: I45914  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-1170 <KAM>  
 A;Cross-References: UNIPROT:P53710; GB:L25886; NID:9439695; PIDN:AB59255.1; PID:9439696  
 F;161-336/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 28.7%; Score 1754; DB 2; Length 1170;  
 Best Local Similarity 33.9%; Pred. No. 4.7e-122; Indels 102; Gaps 30;  
 Matches 411; Conservative 235; Mismatches 455;

Qy 9 LFLPLVFLTGL---CSPFNLDHHPRLFPQPPAEFGYSVLQHVGGQRMWLVGAPWDGP 65  
 Db 2 LQLVLVPSQILNCVAYNVGLPKAKIFSGPSSBQFGYAVQOFINPKGNMLLVGSPWSGF 61  
 Qy 66 SGDRRGDYYRCVPGGAHNAACAAGHL-GDYQLGNSSHPAVNMHGLMSLLETGDDGGPMAC 124  
 Db 62 PKRMGDVYKCPV-DLSTTTCCKLNLTSTSMGNVTETKMTNMSLGLTLTRNVGTGGFLTC 120  
 Qy 125 APLWSRACGSVFSSGICARVDASFQPGSLAPTAQRCPTVMYDVIWLDGNSIYPMSEV 184  
 Db 121 GPLWAQCGSQYTTGVCSDVSPDFQLRTSAPAVQTCPSFIDVVVVCDENSIYPMWAV 180  
 Qy 185 QTFRLRLVGLKFLIDPEQIQVGLVQYGESPVHESLGDFTKBEVRAAKNLSRRBRET 244  
 Db 181 KNFLKPFVQGLDGTGTTQGLIYANNPRVFNLTFSKDEMIKATSTQFYGGDLTN 240  
 Qy 245 TAQIMVACTEGSSQSGHGRPEARLLVVVTDGSHDGBELPAALKACAGRVTRYGI 304  
 Db 241 TPKAIQIARDTAYSTAAGRGPGATKVMVVVTDGSHDGSKLKAVIDQCNKDNLIRFGIAV 300  
 Qy 305 LGHYLRQRDPSPFLRIRTIASDPDRFFNFVTDAAALTDIVDALGDRIFGLEGSHAEN 364  
 Db 301 LGYLNRAALDTKLIKIKAIASITPETHFNFVSDEADLLEKAGTIGEQIFSTEGT-VQG 359  
 Qy 365 ESFGLEMSQIGFSTHRLKDG--ILFGMVGAYDWGGSVLW-LEGHRLFPFRMALEDFP 421  
 Db 360 GDNFQEMSGVGSAAEYSPQNNILMLGAVGAYDWSGTVMQKTPHGLHIFS-----KQAFE 414  
 Qy 422 PALQ--NHAAYLIGYSVSMULLRGRRFLSGAPRFRHGKVIAPQLKKDGAVRVAQSLQ 479  
 Db 415 QILQDRNHSSYLGYSVAS-ISTGNSHFVAGAPRANTYTGQIVLYSVNENGNTVVISQ 473  
 Qy 480 EQIGSYFGSLCLPDTDRDGTDLVLAAPMFLGPQNKETGRVYVYLWGQSLTLTGTL 539  
 Db 474 DQIGSYFGSVLCAVDVKNKDTITDVLVGA PMYNDLKEBGRVYLFITKILNWHOFLE 533  
 Qy 540 QPEPPQDARFGFAMGALPDNLNQGFADVA GAPLEDHGHQALYLYHGTQSGVRPHPAQRI 599  
 Db 534 GPNGLENARFGSAI AALSDINMGDFNDVIVGSPLENQNSGAVYIYNGHEGMIRLRYSQKI 593  
 Qy 600 AAASMPHA--LSYFGRSVGRDLDDGDLVDVAAGAAIILSSRPVHLTSLVETPQ 657  
 Db 594 LGSDFAPSSHLQVPGRLDGDGLNGDSITDVSAGAFQVQVQLMSQSIADVSDATPK 653  
 Qy 658 AISVQRDCRRRQGEAVCLTAALCFQVTSRTPGRDHGFYMRFTASLDEWTAG---ARA 713  
 Db 654 KITLLNKAS-----IKLCLFSAKFR-PTNNQNVAVIYNTITDEQFSSRVISRG 704  
 Qy 714 AFGSGQRLSPRRLRLSVGNVTCQLHFHLDTSYLRPVALVTVPALDNTTKPG--PVL 771  
 Db 705 LFKENNERCLQKTMIVSOAQ-RCSEYIIHQEPSDIIISPLNLCMNISLEN---PGTNPAL 760  
 Qy 772 NEGSPSTIQKLVPSKDCGPDNECVTDIVLQVNMIDIGSRKAPVVRGGRKVLVSTLE 831  
 Db 761 EAYSETVKVSIPIPHKDCGDDGVCSIDLVLNV-QQLPATQQQPFIVSNQNKRLTFSVOLK 819

Qy 832 NRKENAYNTSLSIIFSRNLHLASLTTPQRESPI---KVEC-AAPSAHARLCSVGHVPFQTG 887  
 Db 820 NKESAYNTEIVDFSENLFPAWS-----MPVDGTEVTCQIASSQSKSVTCNVGYPALKSK 875  
 Qy 888 AKVTFLLEPFSCSSLLSQVFGKLTASSDSLENGTLQENTAQTSAYIQYEPHLLFSSES 947  
 Db 876 QQVTFITNFDNLQNLQNASISFRALSSESQEN--MADNSVNLKLSLLYDASIHITRST 933  
 Qy 948 TLHRYE-----VHPYGTLPVGGPGEFKTTLVQNLGCVVSGSLIISALLPAVAHG 997  
 Db 934 NINFEVSLDGNVSSVHVSFEDI---GPKFISIKV-TTGSVPVSMASVLIHIPQYTKD 988  
 Qy 998 GNYFLSLSQVITNNA---SCIVQNLTEPPQPPVHP-----INPLKIGOTSSSVFSENFHRELNC 1040  
 Db 989 KNPLMVLTVGHTDQAGDISCEAE-----INPLKIGOTSSSVFSENFHRELNC 1039  
 Qy 1041 SNTQCOVVRCHLQOLAKGTEVSVGLLRLVHNEFFRRAKPSKLTVTVTFELGTBEGSVLQL 1100  
 Db 1040 RTASCNIMCWLARDLQVKG EYFLNVSTRWNGTFAASTQTQVLTAAAEIDTYNPQIYVI 1099  
 Qy 1101 TEASRWSSESLEVVQTRP---ILISLWILGSLVGLLLALLVFCILMKLGFEAHK---- 1153  
 Db 1100 EE-----NTVTIPTIMKPEKVEVPTGVIVGSIAGLLLLALLVAILMKLGFPRKRYEK 1155  
 Qy 1154 -KIPBEKREKEL 1165  
 Db 1156 AKNPDETTEL 1168

RESULT 5  
 S44142  
 VLA-2 protein homolog - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004  
 R;Edelman, J.M.; Chan, B.M.; Uniyal, S.; Onodera, H.; Wang, D.Z.; Damjanovich, L.; Latzer, S.  
 submitted to the EMBL Data Library, January 1994  
 A;Description: The mouse VLA-2 homologue supports collagen and laminin adhesion but not  
 A;Reference number: S44142  
 A;Accession: S44142  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-1178 <EDS>  
 A;Cross-References: UNIPROT:Q62469; EMBL:Z29987; NID:9473098; PIDN:CAA82877.1; PID:947309  
 F;159-344/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 28.6%; Score 1744; DB 2; Length 1178;  
 Best Local Similarity 35.3%; Pred. No. 2.6e-121;  
 Matches 420; Conservative 221; Mismatches 479; Indels 70; Gaps 31;

Qy 9 LFLPLVPLT-GL--CSPFNLDHHPRLFPQPPAEFGYSVLQHVGGQRMWLVGAPWDGP 65  
 Db 10 LLLQLMLVQGLNCLAYNVGLPGAKIFSGPSSBQFGYSVQQLTNPQGNMLLVGSPWSGF 69  
 Qy 66 SGDRRGDYYRCVPGGAHNAACAAGHL-GDYQLGNSSHPAVNMHGLMSLLETGDDGGPMAC 124  
 Db 70 PENRMGDVYKCPV-DLPATACEKLNQNSASISNVTEIKTNMSLGLTLTENPGTGFLTC 128  
 Qy 125 APLWSRACGSVFSSGICARVDASFQPGSLAPTAQRCPTVMYDVIWLDGNSIYPMSEV 184  
 Db 129 GPLWAQCGNQYATGTCSDVSPDFQLRTSAPAVQTCPSFIDVVVVCDENSIYPMWAV 188  
 Qy 185 QTFRLRLVGLKFLIDPEQIQVGLVQYGESPVHESLGDFTKBEVRAAKNLSRRBRET 244  
 Db 189 KNFLKPFVGLDITGPKKTKQVALIQYANEPRIIFNLDFTKEDMVAQTSSTROHGGDLTN 248  
 Qy 245 TAQIMVACTEGSSQSGHGRPEARLLVVVTDGSHDGBELPAALKACAGRVTRYGI 304  
 Db 249 TFAIEAFARDYASQTSGGRPAGATKVMVVVTDGSHDGSKLKTVIQCNDDDEILRFGIAV 308  
 Qy 305 LGHYLRQRDPSPFLRIRTIASDPDRFFNFVTDAAALTDIVDALGDRIFGLEGSHAEN 364



Db 309 LGYLNRLNLTNKLKEIKAIASPTPTTYFFNVADENALLEKAGTIGEQIFSTEGT-VQG 367  
Qy 365 ESSFGLMSOIGSTHRL--KOGILFGWGYDVGWGSVLWLEGGHR--LPPPRMALEDEF 420  
Db 368 GDNFQEMAGVGSADYAPQNDILMLGAVGAFDWSGLV-QETSHKPVIFP-----KQAF 421  
Qy 421 PPALQ--NHAAYILGYSSVSMLLGGRRLLFLSGAPRFRHGRKVIAPOLKKDGAIVRAQSLQ 478  
Db 422 DQVLQDRNHSSFLGYSVAALISTEDGVH-FVAGAPRANTYTGQIVLYSVNKGQNTVTIQSHR 480  
Qy 479 GEIQSYFGSELCPDPTDRDGTGTVLLVAAPMFLPQNKETGRVYVVLVQOQSLLTLQGT 538  
Db 481 GDOIGSYFGSVLSDVDKXTITDVLVGAFTYNDLKEEGKGYLFTYIKGILNHQOFL 540  
Qy 539 LOPEPPDARFGFAMGALPDINOGFADNAVAGAPLEDHGHQALYLYHGTQSGVRPHPAQR 598  
Db 541 EGPEGTGNARFGSAIALSINMDGFDNIVGSPVENSGAVVYNGHQGTIRTKYSQK 600  
Qy 599 IAAA--SMPHALSYFGSVLGRDLQDGLDVLVAVAGAGAAIILSRPIVHLTPSLEVP 656  
Db 601 ILGNGAFRRHLOFFGRLSDGYDGLNGDSITDVSIGALQVQIQLWSOSIADVAIEALFTP 660  
Qy 657 QATSVQVQDCRRRQGEAVCLTAALCFQVTSRTGRMDHGFYMRFTASLDWTAGARAADF 716  
Db 661 DKITLKNDAK-----ITLKLCPRAEFPAGQ--NNQVAILFNMTLDAGHSSRVTSR 711  
Qy 717 GSGQRLSPRLR--LSGVNV--TCEQLHFHVDTSYLRPVALVTVPALDNTXPG--PVL 771  
Db 712 GVPRENSERFLQNMVNVQKCEHHSISIQKSDVNVPLDLRVDISLEN--PGTSPAL 768  
Qy 772 NECSPTSIQKLVFSDKDCGPNCEVTDLVLQVNDIRGSKAPFVVRGGRKVLVSTLE 831  
Db 769 EAYSETVKVFSIPPYKECGSDGICISDLILDV--QQLPAIQTSQFIVSNQKRLTFSVILK 827  
Qy 832 NRKENAVNTLSIIFSNLHLASITPQRESPI---KYECAAPSAHARL-CSVGHVPVQTG 887  
Db 828 NRGESANTVVLAEFSENLFASF-----MPVDGTEVTCEVSGSKSVTCDVGPALKSE 883  
Qy 888 AKVTFLLEPFSSLLSVOFGKLTASSDSLSLERNGLTQENTAQTSAYIQYEPHLLFSSES 947  
Db 884 QVTFITNFNLQNLQAAINQAFSESOETNKA--DNSVSLTIPLLYDAELHLTRST 941  
Qy 948 TLHRYEHPVGTLP-----VGPQPEPKTLRQVNLGCYVVSGLIISALLPAVAHGGNYF 1001  
Db 942 NINFEYISSDENAFSVIKVEDIGPKFIFSLKV--TAGSAPVSMALVTIHIPOYTKENPL 1000  
Qy 1002 LSLSOVITNNA--SCI--VQNLTEP--PGPVHPELOHTNRLNGSNTQCVVRCHLG 1053  
Db 1001 LYLFGIQTDAQDISCTABEINPLKPHTAFSVSFENFRHTKELDCRTTSCSNITCWLK 1060  
Qy 1054 QLAGTEVSVGLLRVHNEFFRRAPKSLTVVSTFELGTEEGSVLQLTASRWSESLELV 1113  
Db 1061 DLHMAEYFVNTVTRVNRNTPAASTFTQVLTAAAEIDTHNPQ--LFVIEENAVTIPLMIM 1119  
Qy 1114 VQTPILISWILGSLVGLLLALLVFLWKLGP--AHKKI---PEE 1158  
Db 1120 KPTEKAEVPTGVIIIGSIAGILLLLAMTAGLWKLGFPRQYKMGQNPDE 1169

RESULT 6  
S03308  
cell surface glycoprotein cd11a precursor - human  
N;Alternate names: leukocyte adhesion glycoprotein LFA-1 alpha chain; leukocyte function  
C;Species: Homo sapiens (man)  
C;Date: 28-Feb-1990 #sequence revision 28-Feb-1990 #text\_change 09-Jul-2004  
C;Accession: S03308; A47458; A47565; A48759; S36044  
R;Larson, R.S.; Corbi, A.L.; Berman, L.; Springer, T.  
J. Cell Biol. 108, 703-712, 1989  
A;Title: Primary structure of the leukocyte function-associated molecule-1 alpha subunit  
A;Reference number: S03308; MUID:89139587; PMID:2537322  
A;Accession: S03308  
A;Molecule type: mRNA  
A;Residues: 1-1170 <LAR>

A;Cross-references: UNIPROT:P20701; UNIPROT:Q9UBC8; EMBL:Y00796; NID:g31421; PIDN:CAA68  
A;Note: part of this sequence was confirmed by protein sequencing  
R;Cornwell, R.D.; Gollahon, K.A.; Hackett, D.D.  
Proc. Natl. Acad. Sci. U.S.A. 90, 4221-4225, 1993  
A;Title: Description of the leukocyte function-associated antigen 1 (LFA-1 or CD11a) prc  
A;Reference number: A47458; MUID:93248261; PMID:8097887  
A;Accession: A47458  
A;Molecule type: DNA  
A;Residues: 1-20 <COR>  
A;Note: sequence extracted from NCBI backbone (NCBIN:130862, NCBI:P130863)  
R;Shelley, C.S.; Farokhzad, O.C.; Arnaout, M.A.  
Proc. Natl. Acad. Sci. U.S.A. 90, 5364-5368, 1993  
A;Title: Identification of cell-specific and developmentally regulated nuclear factors t  
A;Reference number: A47565; MUID:93281759; PMID:8099450  
A;Accession: A47565  
A;Molecule type: DNA  
A;Residues: 1-20 <SHE>  
A;Cross-references: GB:M95609  
R;Nueda, A.; Lopez-Cabrera, M.; Vara, A.; Corbi, A.L.  
J. Biol. Chem. 268, 19305-19311, 1993  
A;Title: Characterization of the CD11a (alphaL, LFA-1alpha) integrin gene promoter.  
A;Reference number: A48759; MUID:93374910; PMID:8103515  
A;Accession: A48759  
A;Molecule type: DNA  
A;Residues: 1-20 <NUE>  
A;Cross-references: EMBL:Z22804; NID:g311405; PIDN:CAA80461.1; PID:g311406  
C;Genetics:  
A;Gene: GDB:ITGAL; CD11A  
A;Cross-references: GDB:119757; OMIM:153370  
A;Map position: 16p11.2-16p11.2  
C;Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo  
C;Keywords: cell adhesion; cytoskeleton; glycoprotein; heterodimer; surface antigen; tra  
F;1-25/Domain: signal sequence #status predicted <SIG>  
F;26-1170/Product: leukocyte adhesion glycoprotein LFA-1 alpha chain #status predicted <  
F;154-317/Domain: von Willebrand factor type A repeat homology <WMA2>  
Query Match 18.1%; Score 1105; DB 2; Length 1170;  
Best Local Similarity 29.4%; Pred. No. 1.2e-73;  
Matches 366; Conservative 198; Mismatches 465; Indels 216; Gaps 55;  
Qy 11 LPLVFTLGL-----CSPFNLDHHPRLFPGPPEA--EFGYSVLQHVGGQRMVLVGPWD 63  
Db 9 MAMALLSGFFFPAPASYNLDVVGARSP--SPPRAGRHFYRVLQ--VGNG---VIVGAPGE 63  
Qy 64 GPSGDRRGDVYRCPVCGAHNAPCAKHLGDYQLGNSHSPAVNMHLGMSLLETGDDGFMA 123  
Db 64 GNS---TGSLYQCOSGTGCHLPVT-----LRGSNY---TSKYLGMTLATDPTDGSILA 110  
Qy 124 CAPLWSRACSSVFSFSGICARVDASFO--POGSLAPTAQRC--PTYMDVIVLDGNSIYP- 180  
Db 111 CDFGLSRTCDQNTYLSGLCYLFRQLQGPMLQGRPGFQECIKGNVDLVFLFDGSMSLQPD 170  
Qy 181 -WSEVQTFLLRLVKLPIDPEQIQVGLVQYGESVPVHWSLGDG---RTKEVVRRAAKNLS 236  
Db 171 EFQKILDFMDVMKKL--SNTSQPAAVQFSTSYKTEFDFSDYVVKWDPDALLKHVRML 228  
Qy 237 RREGREFTKTAQAINVACTEGFSQSHGRGPEARALLVYVTVTDGESHGDELPALAKACEAGR 296  
Db 229 LL-----TNTFGAINYVATEVFEELGARPDATKVLIIITDGEATDSGNIDAAKD----- 278  
Qy 297 VTRYGI AVLGHYLRQRDPSFLREIRTIASDDPDERFFNVTDAAALTDIVDALGDRIFG 356  
Db 279 IIRYIIGIGHFQTKSQEET-----LHKFASKPASEFVKILDTPEKLKDLTFELQKIIYV 333  
Qy 357 LEGSHAENESSFGLMSOIGFSTHRLKDGILFGVMVGAYDMGGSVLWLEGGHRLPFRMAL 416  
Db 334 IEGTSKODLTSFNMELSSSGISADLSRGHVVAVGAVGANDWAGGFLDLKADLIQ----- 385  
Qy 417 EDEF-----PPALQNHAAVLYGYSSVSMLLRGRRRLFLSCAPFRHGRKVIAPOLKKDGA-V 471  
Db 386 DDTFIGNLEPTPEVRAGYLGTVTWLPSRQKTSLLASGAPRYQHWGRVLLFQBPQGGGHW 445  
Qy 472 RVAQSLQGEIGSYFGSELCPDPTDRDGTGTVLLVAAPMFLGPQNKETGRVYVY---LVG 528

Db 446 SQVQTHGTQIGSYFGELCGVDVDDQGETELLIGAPFYGEQ--RGGRVFIYQRRLG 503  
Qy 529 QOSLLTLQGTLOPEPPQDARFGPAMGALPDLDNDGADVAGAPLEDHGHGALVYHGTO 588  
Db 504 FEEVSELQG--DGOYPL-GRAGEAITALTIDINGDLVDVAVGAPLEB--QAVYIFNGRH 558  
Qy 589 SGVRPHPAQIAAASMPHALSYFGRSVDRGLDLDDGDLVDVAVGAQAAILLSRPVHL 648  
Db 559 GGLSPQPSQRIEQTQVLSGQWFGRSIHGKDLGDLGADVAVGAESQMIVLSRPPVDM 618  
Qy 649 TPSLEVTPOAISVVQDRC---RRRQGEAVCLTAALCFQVTSRTPGRWDHOFYMR----- 699  
Db 619 VTLMSFSPAETIPVHEVECSYSTSNKKEGVNIT--ICFQIKSLYP-----QFQRLVANL 671  
Qy 700 -FTASLDEWTAGARAAADGSGQRLSPRLRLSVGNVTCEQHLHFHV-LDTSYLRPVALTV 757  
Db 672 TYTLQDLGHTRRRGLFPQGRHEL--RRNIATVTSMSCTDFSPHPVVCVQDLISPINVSL 729  
Qy 758 TFAL---DNTTKGFPVLNBSPTSIQKL-----VPSKDCGPDNCEVTDLVQVN-MD 806  
Db 730 NPSLWEEGTPRDQRAQKQIDPILRPSLHSETWEIPEKNCGEDKKCEANLRVSPSPAR 789  
Qy 807 IRGRKAPFVVRGRRKVLVSTTLNKRKENAYNTSLSIIFSRNLHLASLTPQR-ESPIKV 865  
Db 790 SRALRLTAFAP-----SLSVLSLSNLEEDAYVQDLHPFPGLSFRKVEMLKPHSQIPV 843  
Qy 866 ECAAPSAHARL-----CSVGHVPFQNGAKVTFLLBEPFSCSSILSQVFGKLTASS--DS 917  
Db 844 SCBELPEERLLSRALSCNVSSPIFKAGHSA-----LQMFNTLVNSSWGDS 891  
Qy 918 LERNGT-----LQNTAQTSAIYQIEPHLLF--SSESTLHRYEVHPYGTLPVGP 965  
Db 892 VELHANVTCNEDSLEDNSATTIIPILYPINILIQEDSTLY-----VSFTPKGP- 944  
Qy 966 PEFTTLRVNLCGVVSGLIISALLPAVAGHGYFLSLSQVITNNASCIVQNITEPP-- 1023  
Db 945 -----KIHQKHYQVRIQPSI-HDHN-----IPTLEA---VVGVPQPPSE 981  
Qy 1024 GP-----PVHPELQHTNRLNSNTQC---QVVRCHLQGLAQKTEVSUGLLR 1067  
Db 982 GPITHQMSVQMEPPVPCYEDLE---RLPDAAEPCLPALFRCPVPRQELVQVIGTLE 1038  
Qy 1068 LV-----HNEPFRRAKSLTAVST--FELGTGEGSVLQLTEASRWSLELVVQTRPILI 1121  
Db 1039 LVGEIEASSNFCSSLSISFNSSKPHLYGSNASLAQVV-----MKVDVYVVKQMLY 1091  
Qy 1122 SLMTLIGSVLGGLLLLALLVFCMLKQGFPAHKKIPEBEKREKLE 1166  
Db 1092 -LVVLSG--IGGLLLLLLIFVLYKVGFF-----KRNLEKME 1126

RESULT 7  
I56126  
lymphocyte fuction-associated molecule-1-alpha - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 09-Jul-2004  
C/Accession: I56126  
R/Kaufmann, Y.; Tseng, E.; Springer, T.A.  
J. Immunol. 147, 369-374, 1991  
A/Title: Cloning of the murine lymphocyte function-associated molecule-1 alpha-subunit  
A/Reference number: I56126; MUID:91268576; PMID:2051027  
A/Accession: I56126  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-1163 <RES>  
A/Cross-references: UNIPROT:P24063; GB:M60778; NID:g198785; PIDN:AAA39426.1; PID:g198786  
C/Genetics:  
A/Suprafamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat hom  
F/151-315/Domain: von Willebrand factor type A repeat homology <WVAL>  
Query Match 18.0%; Score 1097; DB 2; Length 1163;

Best Local Similarity 28.4%; Pred. No. 4.8e-73;  
Matches 356; Conservative 192; Mismatches 470; Indels 234; Gaps 48;  
Qy 12 PLVFLTGL-----CSPENLDHEHPRLFPQGPPEAFGVSVLQHVGGQRMMLVGPAPMDGPS 66  
Db 8 PRULLLGLQLFAKAWSYNLDTRTQSFQAQGRHFGYQLQIEDG---VVVGAPEB--- 60  
Qy 67 GDRRGDVYRCPVGAHNAPCAKGHLGDYQIAGNSHSPAVNMHLGMSLLETDGCGPMACAP 126  
Db 61 GDNLTGGLYHCRTSSSEFCQPVSV-----LHGSNH--TSKYLGMTLATDAAKGSLAACDP 110  
Qy 127 LMSRACSSVFSGGICARVDASFPQGSGLAPTAQRCTPY-----MDVVIVLDGNSI- 178  
Db 111 GLSRTCDONTYLSGLC-----YLFQSLGPMQLQNPAYQECMKGVLDLVFLFDGSQLD 165  
Qy 179 -YPMSEVQVFLRLVGLKFLIDPEQIOVGIVQVYGESPVHEMSLGDERTKEBVRVAAAKLSR 237  
Db 166 RKDFEKILEFPMKDVMRKL--SNTSYQFAAVQFSTDCRTEFTFLDY-----VKQNKPDV 217  
Qy 238 REGRE-----TKTAQAIMVACTEGFSQSHGRPEAARLLVVVTDGESHGDBELPAALKA 291  
Db 218 LLGSVQPMFLTLTNPAINVVAHVFKESGARPDATKVLVITDGEASDKGNISA--- 273  
Qy 292 CEAGRVTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDRPFNFVTDEAALTDIIVDALG 351  
Db 274 --AHDITRYIIGIKHFVSQKQ-----KTLHTIFASEPVEEFVKILDTFEKLDLEFTDLQ 326  
Qy 352 DRIFGLEGSHAENESSFGLEMSQIGPSTHRLKDGILPGMVGAYDMGSSVLMWLSGGHRLFP 411  
Db 327 RRIYALTEGTRNODLTSPNMESSGISAADLSKGHVVAVGAKDWAGGFJDLR----- 379  
Qy 412 PRMALED-----BFPALQNHAAAYLYGYSVSSMLRGRRRLFLSGAPFRHRGKVIAF 463  
Db 380 -----EDLOGATVQGEPLTSDVRGGLYGVYVAMWTSRSSRPLAAGAPRYQHVQVLLF 434  
Qy 464 QL-KDQAVRVAQSLQGEQIGSYFGSELCPDLTDRDGTDTDLVLAAPMFLGPONKETGRV 522  
Db 435 QAPEAGRWNTQKIEGTQGTGSYFGSELCSVDLDQDGEAELLIGAPLFFGEQ--RGGRV 492  
Qy 523 YVYLVQOSSLTLQGTLOPEPPQD--ARFGPAMGALPDLDNDGADVAVGAPLEDHGHGAL 581  
Db 493 FTV-QRRQSLEFVMSVSELQDGPVPLRFGAATALTIDINGDLVDVAVGAPLEB--QAV 549  
Qy 582 YLYHGTQSGVRHPAQAIAAASMPHALSYFGRSVDRGLDLDDGDLVDVAVGAQAAILLS 641  
Db 550 YIENGKPGGLSPQSQRIQGAQVFGIRWFGRSIHGKDLGDLGDLADVVVGAESGRVVVLS 609  
Qy 642 SRDVLHTPSLEVTPOAISVVQDRC---RRRQGEAVCLTAALCFQVTSRTPGRWDHOFY 697  
Db 610 SRPVDVVTLESPBEIPVHEVECSYSAREEQKHGKVLKA--CFRIKPLTP-----QFQ 662  
Qy 698 MR-----FTASLDEWTAGARAAADGSGQRLSPRLRLSVGNV-----TCEQLHFHV-L 744  
Db 663 GRLLANLSTYLDGHRMRSRGLFPDGSHEL-----GNTSITPDKSCLDGFHFHFPPI 714  
Qy 745 DTSYLRPVALTVTFALDNTTKGFPVLNBSPT-----SIQKLVFPFSKD 788  
Db 715 CIQDLISPINVSLNFSL-----LEEGTDPDQGRAMQPILRPSIHTVTKIEPEKN 766  
Qy 789 CGPDNCEVTDLVQVNMDIRGSKAPFVVRGGRKRVLSTTLNKRKENAYNTSLSIIFSR 848  
Db 767 CGEDKCKEANLTLS-----SPARSGPLRLMSS--ASLAVETLNSGDEAYVWRDLDDPFR 820  
Qy 849 NTHLASLTP-QRESPIKVECAAPSAHARL-----CSVGHVPFQNGAKVTFLLBEPFSCS 901  
Db 821 GLSFRVEMQLQHSRMPVSCSEELTEGSSLLTKLKNVSSSPIKAGQEVSV----- 870  
Qy 902 SLLSQVFGKLTASS--DSLERNQTL-----QNTAQTSAIYQIEPHLLFSSESTL 949  
Db 871 --LOVMFNTLVNNSWEDFVELNGCTVHCENENSSLQEDNSAATHIPLVYPVNLTKEDEN- 927  
Qy 950 HRYEVHPYGTLPVGPGBEPFKTLTVQNVL-----GCYVVVSGLIISALLPAVAGHGYFL 1002

Db 928 -----STLYISFTPKGPKTQVQHVYQVRIQPSAYVDHNMPTLEALVGPVPHSDDL 979

Qy 1003 SLSOVITNN--ASCIVQNLTEPPGPPVHPEELQHTNRLNGSNTQCVV-RCHLQQLAKGT 1059

Db 980 TYTMSVQTDPLVTCSEDLKRPSSAEQ-----CLPGVQFPCPIVFRWEILIQVTGT 1032

Qy 1060 EVSVGLRLVHNEFFRAKPKS-LTVVSTPELGTGEGSVQLQTEASRWSES----LLEV 1114

Db 1033 V-----ELSKIKASSTLSLSSSVSNSSKHFL-VGSKASAEQVLUVKVDLI 1080

Qy 1115 QTRFILSLMILGSLVGLLLALLVFLCMKLGFFFAHKIPBEKEEKL 1166

Db 1081 HEKEML-HVYVLSG--IGGLVLLFLFLALYKVGFF-----KRLKEXE 1122

RESULT 8

RWU1B

cell surface glycoprotein CD11b precursor [validated] - human

N;Alternate names: complement receptor type 3 alpha chain; leukocyte adhesion protein Mac-1; leukocyte integrin alpha chain; neutrophil adherence receptor alphaM chain

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 09-Jul-2004

C;Accession: A31108; A28915; A41600; A30892; A32218; A46526; A26091; 152567

R;Corbi, A.L.; Kishimoto, T.K.; Miller, L.J.; Springer, T.A.

J. Biol. Chem. 263, 12403-12411, 1988

A;Title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11B)

A;Reference number: A31108; MUID:88315033; PMID:2457584

A;Accession: A31108

A;Molecule type: mRNA

A;Residues: 1-1153 <COR>

A;Cross-references: UNIPROT:P11215; GB:J03925; NID:G187284; PIDN:AAA5944.1; PID:G307148

A;Note: part of this sequence was confirmed by protein sequencing

R;Arnaout, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.

J. Cell Biol. 106, 2153-2158, 1988

A;Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor Mac-1

A;Reference number: A28915; MUID:88257215; PMID:2454931

A;Accession: A28915

A;Molecule type: mRNA

A;Residues: 1-499,501-965,'P',967-1153 <ARN>

A;Cross-references: GB:M18044; GB:J03270; GB:M19664; GB:X07421; NID:G186935; PIDN:AAA594

A;Note: the authors translated the codon TAC for residue 1129 as Thr

A;Note: part of this sequence, including the amino end of the mature protein, was confirmed by protein sequencing

R;Shelley, C.S.; Arnaout, M.A.

Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991

A;Title: The promoter of the CD11b gene directs myeloid-specific and developmentally regulated expression

A;Reference number: A41600; MUID:92073318; PMID:1683702

A;Accession: A41600

A;Molecule type: DNA

A;Residues: 1-9 <SHE>

A;Cross-references: GB:M76724; NID:G180018; PIDN:AAA58410.1; PID:G553215

R;Arnaout, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Harris, P.; Tenen, D.G.

Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988

A;Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesion molecule-1

A;Reference number: A94193; MUID:88190151; PMID:2833753

A;Accession: A30892

A;Molecule type: mRNA

A;Residues: 917-1042 <AR2>

A;Cross-references: GB:M18044

R;Hickstein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.M.; Back, A.L.; Roth, G.J.

Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989

A;Title: cDNA sequence for the alphaM subunit of the human neutrophil adherence receptor

A;Reference number: A32218; MUID:89098893; PMID:2563162

A;Accession: A32218

A;Molecule type: mRNA

A;Residues: 9-1153 <HIC>

A;Cross-references: NID:G189068; PIDN:AAA59903.1; PID:G386975

A;Note: part of this sequence was confirmed by protein sequencing

R;Fleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.P.; Tenen, D.G.

J. Immunol. 150, 480-490, 1993

A;Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-1B chain

A;Reference number: A46526; MUID:93123748; PMID:8419480

A;Accession: A46526

A;Status: not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-499,501-1153 <FILE>

A;Cross-references: GB:S52227; NID:G263047; PIDN:AA24821.1; PID:G263049

A;Note: the last three bases of intron 13, CAG, are included in some but not all mature transcripts

A;Note: sequence extracted from NCBI backbone (NCBI:P:121963)

R;Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnaout, M.A.

Biochim. Biophys. Acta 874, 368-371, 1986

A;Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across species

A;Reference number: A90664; MUID:87076671; PMID:3539202

A;Accession: A26091

A;Molecule type: protein

A;Residues: 17-31 <PIE>

A;Experimental source: granulocytes

R;Pahl, H.L.; Rosmarin, A.G.; Tenen, D.G.

Blood 79, 865-870, 1992

A;Title: Characterization of the myeloid-specific CD11b promoter.

A;Reference number: 152567; MUID:92144986; PMID:1346576

A;Accession: 152567

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-9 <RES>

A;Cross-references: GB:M84477; NID:G180184; PIDN:AAA51960.1; PID:G553219

C;Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1

C;Genetics: GDB:ITGAM; CR3A

A;Gene: GDB:ITGAM; CR3A

A;Cross-references: GDB:120599; OMIM:120980

A;Map position: 16p11.2-16p11.2

A;Note: promoter contains a GATA motif and two Sp1 consensus binding sites

C;Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homology

C;Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer; mag-

F;1-16/Domain: signal sequence #status predicted <SIG>

F;17-1153/Product: cell surface glycoprotein CD11b #status experimental <MAT>

F;17-1108/Domain: extracellular #status predicted <EXT>

F;148-318/Domain: von Willebrand factor type A repeat homology <VWA>

F;455-473/Region: calcium/magnesium binding #status predicted

F;530-538/Region: calcium/magnesium binding #status predicted

F;593-601/Region: calcium/magnesium binding #status predicted

F;1109-1134/Domain: transmembrane #status predicted <TM>

F;1135-1153/Domain: intracellular #status predicted <INT>

F;86,240,391,469,693,697,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Binding

Query Match 17.8%; Score 1085.5; DB 1; Length 1153;

Best Local Similarity 29.6%; Pred. No. 3.4e-72;

Matches 366; Conservative 196; Mismatches 492; Indels 183; Gaps 47;

Qy 11 LPLVFLTG--LCSPFNLDHHPRLFCPPAEFGYSVLQHVGGQRMWLCAPWDGSGD 68

Db 3 LRVLLLTALFLCHGFNLDNTENAMTFQENARG-FGQSVVQLQSGR---VVVGAPOEIVAA 58

Qy 69 RRGDVYRCVGGAHNAPCAKHLGDLVQLGNSSHP-----AVNMHLGMSLLETGDGG 120

Db 59 QRGSLVQC-----DYSTG-SCFPIRLQVPVEAVNMVSLGLSLAATTSPQ 101

Qy 121 FMACAPLWSRACSSVFPSSGICARVDASFOQSSLAQAOR-CPTY-MDVVIVLDGNSI 178

Db 102 LLACGPTVHTQCSNTYVKGCLFLGSLNLRQPKFPEALRGCPQEDSIAFLIDGSGSI 161

Qy 179 YPNSVQVTFRLRLVGLKFLIDP--EQIQ-----VGLVQYGESPVHWSLGFRTKEEVVRA 231

Db 162 IPHD-----PRRM--KEFVSTVMEQLKSKTLFLSMQYSEEFRIHFTFKFQNNPNRSL 214

Qy 232 AKNLSRRREGRETQTAQAIMVACTEGFSQSHGSPPEARLLVWTDGSHDDEELPAALKA 291

Db 215 VKPIQLLGR-THATGIRKVVRELFNITNGARKNAFKILVITDGEKF-GDPLGYEDVI 272

Qy 292 CEAGR--VTRYGIATVGLHYLRQRDPSSFLUREIITIASDPDERFFNVVTDEAALTIDVA 349

Db 273 PEADREGVIRYVIGV-GDAFRSEKS-----RQELNTIASKPPDRHVFQVNNFEALKTIONQ 327

Qy 350 LGDRIIFGLESHAENSSFGLENSQGFSTHRLKDGILFGMVGAYDWGGSVLMLEGGHRL 409

Db 328 LREKIFAEGTGTGSSSFHEMSQEGFAAITSNGPLLTSTVSGYDWAGGVLYTSKEKS 387  
Qy 410 FPRPMALDEFPFPPALQNHAAAYLYGSVSMLLRGRRLLFLSGAPFRHGRKVIAPOLKKDG 469  
Db 388 TFINMTVDVS-----DNMDAYLVGA-AAILRNVRQSLVLGAPRYQHIGLVAMFR-QNTG 440  
Qy 470 AVRVAQSLQEQIGSYFGSELCPDTRDGTDTVLLVAAPMFGLPQNKETGRVYVYL-- 527  
Db 441 MWESNANVKGQTGIGAYFGASLCSVDVDSNGSTDLVLIGAPHYY--EQTRGGQVSVCLPR 498  
Qy 528 GQSSLLTLOGTLQPEPPD-ARFGFANGALPDINQDGFADVAVCAPLEDHGHQALYLYHG 586  
Db 499 GORARMCQDAVLYGEOQPMGRFGAALTVLGDVNGDKLTDVAIGAPGEENRGAFLYFHG 558  
Qy 587 TQ-SGVAPHAPQRIAAASMPHALSYFGKSYVDGRLLDDGDDLDVAVGAQAAILLSRPI 645  
Db 559 TSGSGISPSHSQRIAGSLSPRLQYFGQSLGGQDLTMDGLVDTVGAQGHVLLRSQPV 618  
Qy 646 VHLTPSLEVTFQALISVVQRDCCR---RQGEAVCLTAALCFQVTSRTFGRWDH---QPYMR 699  
Db 619 LRVKAIMFNPREVARNVFECDQVVKGEAG--EVRVCLHVQKSTRDLREGQIQSVVT 676  
Qy 700 FTASLDWTAGARAAFGSGORLSPRRLRSVG-NVTCEQLHFHVLDP-TSDYLRLPVALT 757  
Db 677 YDLALDSGRPHSRAPFNET--KNSTRQTVGLGTQCTELKQLPNCIEDPVSPVLR 734  
Qy 758 TFDALDNTTKPG-----PVLNEGSPSTOKLVPFSKCGPNECVTDLVLO---VNMDIRG 809  
Db 735 NFSLVGTPLSAFGNLRPVLAEDAQRLLTALFPPEKNGCNINCOODLSIFSPMSLDC-- 792  
Qy 810 SRKAPFVVRGRRKVLVSTTLENKENAYNTSLIIFSRNL---HLASLTPQR-ESPIKV 865  
Db 793 -----LVVGGPREFNVTVIRNDGEDSYRTQVTFPPFLDLSYRKVSTLQNRQSRWRL 846  
Qy 866 ECAAPSA-----HARLCSVGHVPFQTGAKVFTLLPEFSPSCSLLSQVQKLTASDLS 918  
Db 847 ACESASSTEVSGALKSTCSINHPFENSEVETNITFDVDSKASLG---NKLK-----L 898  
Qy 919 ERNGLTQENTAQ-----SAYIQYBPH-----LLPSSSESTLHRYBVHPYGT 959  
Db 899 KANVTSENMPRTNKTBEQLELPVKYAVYVVTVSHGVSTKYLNFTAGENTSRYVMQHY-- 956  
Qy 960 LPVGPGEPEKTKTLRVQNLGCVVWSGLIISA--LLPAVAHGNGYFLSLSQVI----- 1008  
Db 957 -----QVSNLG---QRSPLSLVFLVP-----VRLNQTVIMDRPQVTF 991  
Qy 1009 TNNAACVQNLTEPPGPPVHPELOHTNRLNGSTQCVVRCHLQGLAKGTEVSVGLRL 1068  
Db 992 SENLSSTCHTKERLPHSDFLAELRKAPVNVNCSIAVCQRIQCDDIPFFGIQBEFNATLGN 1051  
Qy 1069 VHNFFPRRAKSLTVVSTPELGTREGSVLQLTEASRWSLLEV-VQTRPILISLILI 1127  
Db 1052 LSPDWYIKTSHNHLIVSTAEILNDSVFTLLPQGGAFVRSQTKVPEPVPNPLIV 1111  
Qy 1128 GSVLGGLLALLVFLKWLKGFPAHKKIPBEEKREEK 1164  
Db 1112 GSVVGLLLALLAALVYKLGFF-----KQYK 1139

## RESULT 9

## RWHLIC

Cell surface glycoprotein CD11c precursor - human

N/Alternate names: leukocyte adhesion receptor p150,95 alpha chain

C/Species: Homo sapiens (man)

C/Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 09-Jul-2004

C/Accession: A36584; A35543; S00864

R/Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.

J. Biol. Chem. 265, 12750-12751, 1990

A/Reference number: A36584

A/Contents: erratum

A/Accession: A36584

A/Molecule type: DNA

A/Residues: 1-1163 &lt;COR&gt;

A/Cross-references: UNIPROT:P20702  
R/Note: this revision to the sequence from reference A35543 includes the carboxyl end  
A/Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.  
J. Biol. Chem. 265, 2782-2788, 1990  
A/Title: Genomic structure of an integrin alpha subunit, the leukocyte p150,95 molecule.  
A/Reference number: A35543; MUID:9015306; PMID:2303426  
A/Accession: A35543  
A/Molecule type: DNA  
A/Residues: 1-834 <CO2>  
A/Note: this sequence has been revised in reference A36584  
R/Corbi, A.L.; Miller, L.J.; O'Connor, K.; Larson, R.S.; Springer, T.A.  
EMBO J. 6, 4023-4028, 1987  
A/Title: cDNA cloning and complete primary structure of the alpha subunit of a leukocyte  
A/Reference number: S00864; MUID:8816645; PMID:3327687  
A/Accession: S00864  
A/Molecule type: mRNA  
A/Residues: 1-755, 'L', 757-1163 <CO3>  
A/Cross-references: GB:M81695; EMBL:X00093; NID:G487829; PIDN:AA59180.1; PID:G487830  
A/Note: part of this sequence was confirmed by protein sequencing  
C/Comment: A common beta chain (CD18) forms a heterodimer with CD11b to form Mac-1 on myelomonocytic cells.  
C/Genetics:  
A/Gene: GDB:ITGAX; CD11C  
A/Cross-references: GDB:119758; OMIM:151510  
A/Map position: 16p11.2-16p11.2  
C/Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homology  
C/Keywords: calcium; cell adhesion; glycoprotein; heterodimer; magnesium; tandem repeat;  
P:1-19/Domain: signal sequence #status predicted <SIG>  
P:20-1163/Product: cell surface glycoprotein CD11c #status predicted <MAT>  
P:20-1107/Domain: extracellular #status predicted <EXT>  
P:149-319/Domain: von Willebrand factor type A repeat homology <VWA4>  
P:1108-1133/Domain: transmembrane #status predicted <TM>  
P:1134-1163/Domain: intracellular #status predicted <INT>  
P:61,89,392,697,735,899,939,1050/Binding site: carboxydrate (Asn) (covalent) #status predicted <AS>

Query Match 17.3%; Score 1054; DB 1; Length 1163;

Best Local Similarity 28.5%; Pred. No. 7.8e-70;

Matches 356; Conservative 201; Mismatches 474; Indels 220; Gaps 50;

Qy 13 LVFLTLGLCSF--FNLDEHHPRLFPPEPEAFGYSVLQHVGGQRMVLVGAPEMDGPDGDRR 70

Db 8 LLLFTALATSLGNLTDELTAF-RVDSAGFGSVVQVYANS---WVVGAPQKITAANQT 63

Qy 71 GDVYRCFVGGAHNAPCAKGLHDYQLGNSHP--AVNMHLGMSLLETGDDGGMACAPLWS 129

Db 64 GGLYQC-----PIGLQVPPAVNMVSLGLSLASTTSPSOLLACGPTVH 113

Qy 130 RACGSSVFSFGICARVDASFPQSGSLAPTAQRPTY-MDVVIVLDGNSIYP--WSEVQT 186

Db 114 HECGRNMYTLGLCLLGLPT-QLTQRLPVSQECPCQEQDIIVFLIDGSGSISSRNFAFMN 172

Qy 187 FLRLVVKLFIDPEQIQVGLVQYGESPVHWSLGDPTKEEVVRAAKNLSRRRGRETAKTA 246

Db 173 FVRAVLSQ-FQRP-STQFSLMQSNKQTHTEFRTSNPLSLASVHQLQG-FYTTA 229

Qy 247 QAIMVACTEQFSQSHGSRPEARLLVVVTDGESH-DEEELPAALKACEAGRVTRYGIATL 305

Db 230 TAIQNVVHRLPHASYGARRDATKILIVITDGKSGSLDYKVI PMADAAGIIRYAIQVG 289

Qy 306 CHYLRRQDPSSFLERITIASDPDRFPFNVVDEAALTDIVDALGDRIRGLSGSHANE 365

Db 290 LAFQNR-----NSWKELNDIASKPSQEHIPKVEDFDALQIDQIQOLKEKI FAIGTETTS 344

Qy 366 SSFGLMSQIGFTHRLKDGILPGMVGAYDVGWSVLWLEGGHRLFPFPMALDEBFPFALQ 425

Db 345 SSFELEMAQEGFAVTPDQGVLAGVGSFTW-----SCGAFLYPNMS-----PTPI 391

Qy 426 NHA-----AYLGYSVSMILRGRRRLFLSGAPFRHGRKVIAP-QLKKGAVRVAQS 476

Db 392 NMSQENVMDRDSYLGYSTELALNKGQSLVL-GAPRYQHTGKAVIFQVSRQW--RMKAE 448

Qy 477 LQGEQIGSYFGSELCPDTRDGTDTVLLVAAPMFGLPQNKETGRVYV----- 524

Db 449 VTGTQIGSYFGASLCSVDVDTDGTDLVLIGAPHYY--EQTRGGQVSVCLPRGWRWMC 506

```
QY 525 --YLVGQSLTLTQCTLOPEPPQDARFGFAMGALPDNLNODGFADVAVGAPLEDGCGALY 582
Db 507 DAVLYGEG-----HPW-----GRFGAULTVGDVNGDKLTDDVVGAPGEEENRGAVY 554
QY 593 LYHGTQS--GVRPHPAQRIAAAMPHALSFGRSVDGRDLDDGLDLDVAVGAQGAAILLS 641
Db 555 LFHGVLPSPISFSPHSQRIAGSQLSRRLQYFGQALLSGGDLTQDGLVDLAVGARGOVLLR 614
QY 642 SRPIVHLTPSLEVTPQAISVVQDRCRR--GQAVCLTRALCFQVTSRTP-----GRWDHQFY 697
Db 615 TRPLWVGVSQMFPAEIPRSFAFCEQVSEQLVQSNICLYIDKRSKNLLGSRDLQSS 674
QY 698 MRFTASLDWTAGARAAFGSGQRSLSPRLRLSVGNVTCEQLHEHVLDTs---DYLRPVA 754
Db 675 VTLDLALDPGLSPRAFTQETNR--SLSRVVLGLKAHCE--NFNLLLPSCVEDSVTPIT 731
QY 755 LTVTFALDNTTKP-----GPIVNEGSPSTSIQKLVFPKDCGPDNECVTDLVLVQVNDI 807
Db 732 LRLNFTL--VGKPLAFRNLRPLMAADAQRYFTASLPEKNCAGADHICQDNLGISFSPF-- 788
QY 808 RGSRAKPFVVGGRKVLVSTTL-----NRKENAYNTSLSIIFS-----847
Db 789 -----GLKSLLVGNLELNAEVMVMWMDGDSYGT--TITFSPHAGLSRYRYVAE 834
QY 848 -----RNLHLASLTQRESPIKVECAAPSAHARLCSVGHVPFQTGAKVTFLEFEFSC 900
Db 835 GKQGLRSLHLTC-----DSAPV-----GSGQTWSTSCRINHLIRFGGAQITFLATFDVSP 886
QY 901 SLLSQVFGKLTASSDLSLRNGTLQENTAQ-----TSAYIQYEPHLLFS---944
Db 887 KAVLGDLRL--LTANVSSENNTPTSKTYFQLELPKVAIVTVSSHQFTKYNLFSESEE 945
QY 945 --SESTLHREYVHYG--TLVPGGPEFKTTLRVQNLGCVYVSGLIISALL--PAVAHG 998
Db 946 KESHVAMHRYQVNNLGORDLPV-----SINFVVPVELNQEAVMMDVEVSHQP 992
QY 999 NYFLSLQVITNASCIVQNLTPEPPGVHPHEELOHTNR-----LNGSNTQCVQVRCHLQ 1055
Db 993 NPSLRCS-----SEKIAPPA--SDFLAHIQKNPVLDCSTAGCLRPRCDVPSF 1037
QY 1056 AKGTVEVSGLLRLVHNEFFRRAKFKSLTVSTFELTEGSGVLQTEASRWSESLLEVVQ 1115
Db 1038 SVQBELFTLKGNLISFGVHQRIQKQKVSVSVAITETDTSVYSQLPQGEAFMRA-----Q 1092
QY 1116 TRPLISLMI-----LIGSVLGLLLALLLVFLKWLKLPFF--AHKKIPEE 1158
Db 1093 TTTVLEKYKVNPTPLIVSGSIGGLLLALLITAVLYKVGFPKQYKEMEE 1143

RESULT 10
S00551
leukocyte surface glycoprotein Mac-1 alpha chain precursor - mouse
N;Alternate names: complement-3 receptor alpha chain
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: S00551; I59078
R;Pyteila, R.
EMBO J. 7, 1371-1378, 1988
A;Title: Amino acid sequence of the murine Mac-1 alpha chain reveals homology with the i
A;Reference number: S00551; MUID:88312584; PMID:3044779
A;Accession: S00551
A;Molecule type: DNA
A;Residues: 1-1153 <P>T>
A;Cross-references: UNIPROT:P05555; EMBL:X07640; NID:G52982; PID:CAA30479.1; PID:G52983
A;Note: the authors translated the codon CAC for residue 569 as Gln
R;Satre, L.; Roman, J.M.; Teplov, D.B.; Dreyer, W.J.; Gee, C.E.; Larson, R.S.; Roberts,
Proc. Natl. Acad. Sci. U.S.A. 83, 5644-5648, 1986
A;Title: A partial genomic DNA clone for the alpha subunit of the mouse complement recep
A;Reference number: I59078; MUID:86287312; PMID:2942940
A;Accession: I59078
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
```

A;Residues: 11-44 &lt;RES&gt;

A;Cross-references: GB:M14293; NID:G198993; PID:AAA39484.1; PID:G554193

C;Genetics:

A;Gene: Mac-1

C;Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo

C;Keywords: cell adhesion; glycoprotein; transmembrane protein

F;1-16/Domain: signal sequence #status predicted &lt;SIG&gt;

F;17-1153/Product: leukocyte surface glycoprotein Mac-1 alpha chain #status experimental

F;1148-318/Domain: von Willebrand factor type A repeat homology &lt;VWA2&gt;

F;1106-1129/Domain: transmembrane #status predicted &lt;TM&gt;

Query Match 16.7%; Score 1022.5; DB 2; Length 1153;

Best Local Similarity 28.6%; Pred. No. 1.7e-67;

Matches 353; Conservative 218; Mismatches 481; Indels 181; Gaps 50;

QY 13 LVFLTGLCSFNLDEHHPLRFPGPPEAFGYSVLQHVGGQRMWLVGAPMDPGSDRRGD 72

Db 7 LVLTALALCHGFNLDEHPMTFQENAKG-FGQNVVQ-LGGTS--VVVAAPQAKAVNQTA 62

QY 73 VYRCPVGGAHNAPCAKHLGDYQLGNSHP-----AVNMHLGMSLLETDDGGFMAC 124

Db 63 LYQC-----DYST-SRCHPIPLQVPPPEAVNMSLGLSLAVSTVPQQLLAC 105

QY 125 APLMSRACGSVSFSGICARVDAS-FOPQGS LAPTAORCPTY-MDVVILVDGSNSI--YP 180

Db 106 GPTVHQNKENTYVNGLCYLFGSNLLRPPQOFPPEALRECPQESDIYFLIDGSSINID 165

QY 181 WSEVQTFRLRLVKGFLIDPEQIQ-----VGLVQYGESPVHWSLGDFRTKEEVVRAAKNL 235

Db 166 FQMKKEFVTVM-----EQFKSKTLFSLMQYSDEFRIHFTFNDFKRNPSPRSHVSP 218

QY 236 SRREGRTKTAQAIWACTEGFSQSHGCRPEAARLLVVVTDGESHGDEELPAALAKACAG 295

Db 219 KQLNGR--TKTASGIRKVVVRELPHKTNAGARENAKILVITDGEKF-GDPLDYKDVPEAD 276

QY 296 R--VTRYGI AVLGHYLRQRDPSSFLREIRTIASDDPERFFENVTDAAALTDIVDALGDR 353

Db 277 RAGVIRVIGV-GNAFNK---PQS-RRELDITASKPAGEHVQVDNFALNTIQOQEK 331

QY 354 IFGLGSHABNESSFGLEMSQIGFSTHRLKDGILFGVMGAYDW--GGSVLMELEGHRLFP 412

Db 332 IFALGEGTQTSSTSEHEMSQEGFSASITNSGPLGSGVSGDFWAGGAFYTSKDKVFIN 391

QY 413 RMALEDEFPALQNHAAIYGVSVSSMLRGRRILFLSGAPFRHRGKVIAPOLKKDQAVR 472

Db 392 TTRVDSM-----NDAYLGYA-SAVILNRVOSLVLAGAPRYQHIGLVNMF--ENFGTWE 443

QY 473 VAOSLQGEQIGSYFGSELCPDITDRCGTVDVLLVAAPMELGPQNKETGRVYVVLVQ--- 529

Db 444 PHTSIKSGQISYFGASLCSVDMDADGNTNLIIGAPHYY--EKTRGGQVSVCPPLPRGA 501

QY 530 --QSLTLTQCTLOPEPPQDARFGFAMGALPDNLNODGFADVAVGAPLEDGCGALYHGT 587

Db 502 RWQCEALLHGD-QGHP--WGRFGAULTVGDVNGDKLTDAVAGPGEQNGQAVIYGA 558

QY 588 Q-SGVRPHPAQRIAAAMPHALSFGRSVDGRDLDDGLDLDVAVGAQGAAILLSRPV 646

Db 559 SIASLSASHSHRIIGAHFSFGLQVFGSLSGKDLTMDGLMDLAVGAQGHLLLRAPVL 618

QY 647 HLTPSLEVTPQAISVVQDRCRR-----GQAVCLTRALCFQVTSRTPCR---WDHQP 696

Db 619 RLEATMEFSPKKVARSVFACQEVQLKVKDAGEVRVCL-----RVRKNTKDLREGDIQS 672

QY 697 YMEFTASLDWTAGARAAFGSGQRSLSPRLRLSVGNVTCEQLHFLHVLDTSDYLRPVAL 755

Db 673 TVTYDLALDPVRSIRAFDET-KNNTRRRTQVFLGMQKCELTLLIPLDCVDSVSPIIL 731

QY 756 TVTFALDNTTKP-----GPIVNEGSPSTSIQKLVFPKDCGPDNECVTDLVLVQVNDIR 808

Db 732 RLNVTL--VGEPLRSFGNLRPLVAMDAREFTAMFPEPKNGNDISICQDD--LSITMSAM 787

QY 809 GSRKAPFVVGGRKVLVSTTLNRKENAYNTSLSIIFSRLHL-----ASLTQRESPIK 864







Db 120 DGRILACAPRLVLTQSTISPT-----141  
Qy 176 NSIYPMSEVQTFLLRLVGLFIDPQIQVGLVQGVSPVHMSLGDPRTKKEVVRAAKNL 235  
Db 142 -----141  
Qy 236 SRREGRETKTAQIMVACTEGFSQSHGGRPEAARLLVVVTDGSHDGEELPAALAKACEAG 295  
Db 142 -----DKEREPT-----148  
Qy 296 RVTRYGIAVLGHVLRORDSSFLREIRTIASDDPRFFNVNVTDEAALTIDVALGDRIF 355  
Db 149 -----GTCFVGH-----SDFTNFVYSPCQSTORD-----LF 175  
Qy 356 GLEGSHAENESSFGLEMSQIGFSPHRLKDG--ILFGMVGAYDMGGSVLMLEGGHRLPPPR 413  
Db 176 GFD-----KITHCQAGFSAQIPSDNSTLVMGAPGSI-----YLQG--QIFPAQS 216  
Qy 414 MA-LED-----EFPALQNHAAVLGYGVSSMLLRG-QRRFLSGAPRFRH-RGKVIAFOL 465  
Db 217 LSTLSDVSNTPQAVAFDN--SYRGYSALGDFNGDGLDYVVGTPRGESLRLGLVAIFD- 273  
Qy 466 KKGAVRVAOSLQEQIGSYFSGELCPDTRDGTDTVLLVAAPMFL-----GPQNKETR 521  
Db 274 --QSLVEIITPVVGEQIVSYFGYSVASVDVNGDGLDD-LLVGAPMFTNREPATEKWEAGR 330  
Qy 522 VYVYL-----VGOOSLLTLQGTLOPEPPQDARFGFANGALPOLNODGFADVAGPLE- 574  
Db 331 VYVYLNADHSLGAPQMLTKKI-----RARPGFITISGDSNQDGFNDVAIGAIDYG 383  
Qy 575 DGHGALYLYHGTSQVGRPHPAQRIAAASMPHA-LSYFGRSVDGRLLDGDGLVDVAVGA 633  
Db 384 EDNSGVVYVHGSABGLRLTESQVLTSELGFSDIITTFGFSVGGQDMQNDYDPLVGA 443  
Qy 634 QG--AAILLSRPIVHTPSLEVTPQAIISVVQRDCR--RRGOEAVCLTAALCFQVNT-SRTP 689  
Db 444 ESADAAILVRTPVVLLEAEITBPIGINLKNKYELPDGTMVTSFVAMACFTVTGNHLP 503  
Qy 690 GRWDHQFVMTASLDEWTAGARAAPDGSQRLSPRLSLVGNVTCEQLHFHVLDT-SD 748  
Db 504 ARIGISVTLTVDDSSI-----TSGRALLVNELSVQTKNRNLDVGMKFCDDPLRAYVVNTIQD 560  
Qy 749 YLRVALTVTFALDNTT-----KGPVLNBSPTSIOKLVPFSDKCPDNECVTDLVQV 803  
Db 561 KLTPADVQLVETDESILLPYELPIINKEAVSSQTKQVSIQNNC-VNNICIPBIGITV 619  
Qy 804 NMDIRGRKAPFVVRGRRKVLVSTTLLENKENAYNTSLIIFSRNLHLASLTTPQRESPI 863  
Db 620 TPNL-----PNVIGQAQELTLVVSINNREGDAFQSLAVVYPEGLQYVRLERRANMDF 673  
Qy 864 KVECAAPSAHARL-CVSGHPVQFGAKVTPLEPEFSCSLLSOVFGK-----LTASS 915  
Db 674 SVTCTEDSALRMITCDGNPLVG-----KYNLEFGLTLSTL--QVSGDKONIEPYLVAGS 726  
Qy 916 DSLERNGTQONTAQTSAIQYEB--HLLFSSESTLHRYEHPYGTLPVPG-----964  
Db 727 ENNEDPNTLNNELNVAVIVDATKLLSASYPEIVTYRVPEDNIVPEFPTKNASEADI 786  
Qy 965 GPEFKTLRVONLCYVYVGLIISALLPAVAHGNYFLSLISQVITN-NASG-IVQNLTPE 1022  
Db 787 GMEVVHLYEVRNTGSSNAAEVTLNIRWPEKDENDGYLLFYLLGIMTDEGVTQCISQGANP 846  
Qy 1023 PGPPVHPEELQHTNRLNGSNTO-----CQVVRCHLGQL 1055  
Db 847 LGVKL-----EASTKEQLSNSTTVSGRRKRGEYAEALAAQAEPICTPESCVLINCTIDEI 903  
Qy 1056 AKGTEVSVGLRLVHNBFFRRAKFKSLTVVSTFELGTEEGSV-----LQLTBASR 1105  
Db 904 KATSKVRIILGRFWERTFQKAVSEAVPVQIOTLASTATATVRSIPIYNIPLMFEFTDSTK 963  
Qy 1106 WSESL-LEVVQTRPILSLMILIGSLGGLLLALLVFLWKLQGFPAHKKIPESKEEBE 1163

Db 964 ASTLITABELVLPVSVSIAMWIIIVSVLGGIILLIILGLMKGPFERKKPGEDQKEYE 1022  
RESULT 14  
A41131  
lymphocyte-Peyer's patch adhesion molecule alpha 4 precursor - mouse  
N;Alternate names: integrin alpha-4  
C;Species: Mus musculus (house mouse)  
C;Date: 05-Jun-1992 #sequence\_revision 05-Jun-1992 #text\_change 09-Jul-2004  
C;Accession: A41131; S16742  
R;Neuhaus, H.; Hu, M.C.T.; Hemler, M.E.; Takada, Y.; Holzmann, B.; Weissman, I.L.  
J. Cell Biol. 115, 1149-1158, 1991  
A;Title: Cloning and expression of cDNAs for the alpha subunit of the murine lymphocyte-  
A;Reference number: A41131; MUID:92064645; PMID:1840602  
A;Accession: A41131  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-1039 <NU>  
A;Cross-references: UNIPROT:Q00651; EMBL:X53176; NID:G51484; PIDN:CAA37316.1; PID:G51485  
C;Superfamily: integrin alpha-4 chain  
C;Keywords: cytoskeleton; transmembrane protein  
Query Match 11.3%; Score 690; DB 2; Length 1039;  
Best Local Similarity 22.7%; Pred. No. 9.4e-43;  
Matches 286; Conservative 186; Mismatches 396; Indels 392; Gaps 59;  
Qy 23 FNLDHHPRLFCGPPEAEFGYSVLQHVGGQRWMLVGAP---W-DGPGDGRDGDVYRCVP 78  
Db 41 YNLDPENALLYQGPSITLFGYSVLVHSHGSKWLVGAPTASWLSNASVVPQAIYRCGI 100  
Qy 79 -----GGAHNAPCAKGLHDYQLGNSSHPAVNMHLGMSLLETGDP-GGFMAC 124  
Db 101 RKNPNQTCBQLQSGSPSGPCGKTCLEERD-----NOMLVTLRQPGENGSIIVTC 151  
Qy 125 APIWSRACSSVPSSGICARVDASFOQGSPLAPTAORCTYMDVIVLQGSNSIYPSSEV 184  
Db 152 GHRW-----KNIF-----YMKSDNKUPT-----169  
Qy 185 QTFLRLVLGKLFIDPQIQVGLVQGVSPVHMSLGDPRTKKEVVRAAKNLSSREGRETK 244  
Db 170 -----GICYMP-----SDLRTE-----LSKR-----186  
Qy 245 TAOAIMVACTEGFSQSHGGRPEAARLLVVVTDGSHDGEELPAALAKACEAGRVTRYGIAV 304  
Db 187 -----MAPCYKDYTRKFGEN-----FASCQAG-----208  
Qy 305 LGHYLRQRDPSPFLREITIASDPPER-----FFPNVTDEAALTIDVALGDRIFGLE 358  
Db 209 -----ISSFYTDLIVMGAPGSSYMTGTVFVYNIT-----TNQYKAFYDR-----248  
Qy 359 GSHAENESSFGLEMSQIGFSTHRLKDGILFGMVGAYDMGGSVLMLEGGHRLPPRMALED 418  
Db 249 -----QNVQVFG-----255  
Qy 419 EFPALQNHAAVLGYGVSSMLLRGGRRL-FLSGAPRFRHGKVIAPQLKKDQAVRVAQSL 477  
Db 256 -----SYLGYSVGAGHFRSPHTTEVVGAPQHEQIGKAVIFSI-DENELNIVYEM 304  
Qy 478 QGBOIGSYFGSELCPDTRDGTDTVLLVAAPMFLGPQN--KETGRVYVVLVQQS--LL 533  
Db 305 KGKGLASVYFGASVCAVDLNAFGSD-LLVGAPM-----QSTIREGRVFVYVINGMGAMV 359  
Qy 534 TLOGTLQPEPPQDARFGFANGALPDNLNODGFADVAVGAPLEDGHQCALYLYHGTQSGVRP 593  
Db 360 EMBRLVSGDKYAAARFGESIANLGDIDNDGFEDIAFGAQEDDLRGAIVYINRGVGISS 419  
Qy 594 HPAQRIAAASMPHALSYFGRSVDGRLLDGDGLVDVAVGA--QGAAILSSSRPIVHLTPS 651  
Db 420 TYSORIEGQGIKSLRMFGQSIGQIDADNNGVVDVAVGAFQSDSAVLLKTRPWWIVEAS 479  
Qy 652 LEVTPQAIISVVQRDCRRRQGEAVCLTAALCFQVTSR-TPGRWDHQFMYMFTASLDE---- 706  
Db 480 LS-HPESVNRKTPDCTENGLPSVCMHLTLCFSYKGEVPG-----YIVLFLVFNVSLDVHRKA 534

QY 707 -----WTAGARAAFDGSGQRSLRRLRLSVGNVTCEQLH---FHVLDTSVLRPVALT 756  
Db 535 ESESRFFYFNGSDVITGS-----IRVSSGKCR-RTHQAFMRKQVDRDLTPHIVE 585  
QY 757 VTPALDN--TTKPG-----PVLNKGSPS-IQKLVPSKCGPDNECVTDLVQVNM 805  
Db 586 ATYHLGHVITKNTTEFFPLOPILOQKKEKQVIRKMINFARFCAYEN-CSADLQVSAKV 644  
QY 806 D-IRGSRKAPFVVRGGRKVLVSTLTENRKENAYNTSLSIIFSRNLHLASLTTPQRESPIK 864  
Db 645 GFLKPYENKTYLAVGSMKTMVSLNAGDDAYETTLNQLPTGLYFIKILDEEQIN 704  
QY 865 VECAPSAHARL-CSVGHVPFQTC--VTFLEBFESCSLLSQVFGKLTAS-SDSLER 920  
Db 705 CEYTESGIVKACSLGIYVDRSLRIDISFLDV-----SSLSRAHEDLSISVHASCEN 759  
QY 921 NGTL---QENTAQTSAYIQEPHLLFSSESTLHRYEVP---YGTLPVGFGE----- 967  
Db 760 EGELDQVRDRNVLTITPLRYEVM-----TVHGL-VNPTSFVYGSSEENE-PETCMAEK 811  
QY 968 FKTLRQNGVYVSGVGLIISALLPAVAGGNVFL-----SLSQVITNASC----- 1014  
Db 812 LNLTFHVINTGISMAPNVSVKIMVP-----NSFLPQDDKLENVLDVQTTTGOCHPKHYG 865  
QY 1015 -----JVQNLTEPPGPPVHPELOHTNR-----LNGSNVQCQVVRCHLQOLAGTE 1060  
Db 866 RECTFAQKGIAGTLTD-----IVKFLSKTDKRLLYCMKADQCHLDFLNCFKMESGKE 919  
QY 1061 VSVGLLRVHNEFFRRKAKSLTWST--FELGT-----EEGSLQLTEASRSESLE- 1112  
Db 920 AS-----VHIQEGRSIILEMETSLSKFEIKATAPPEHPKVIENLKOENVAHVLEG 973  
QY 1113 VQTRP-----ILISLWILIGSVLGLLLALLVFC-LWKLGF--AHKKIPEEEKREE 1163  
Db 974 LHQRPKRHTIIITISLLG-----LIVLLISCVMMKAGFKFKQYKSIQLEENRD 1027

RESULT 15  
S06046  
Integrin alpha-4 chain precursor - human  
N;Alternate names: CD49d; very late antigen-4 alpha chain; VLA-4 alpha chain  
C;Species: Homo sapiens (man)  
C;Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 09-Jul-2004  
R;Accession: S06046; A39355; D28018  
R;Takada, Y.; Ellices, M.J.; Crouse, C.; Hemler, M.E.  
EMBO J. 8, 1361-1368, 1989  
A;Title: The primary structure of the alpha(4) subunit of VLA-4: homology to other integ  
A;Reference number: S06046; MUID:89356603; PMID:2788572  
A;Accession: S06046  
A;Molecule type: mRNA  
A;Residues: 1-1038 <TAK>  
A;Cross-references: UNIPROT:P13612; GB:X16983; EMBL:X15356; MID:g33945; PIDN:CAA34852.1.  
R;Rosen, G.D.; Birkenmeier, T.M.; Dean, D.C.  
Proc. Natl. Acad. Sci. U.S.A. 88, 4094-4098, 1991  
A;Title: Characterization of the alpha-4 integrin gene promoter.  
A;Reference number: A39355; MUID:91239513; PMID:2034655  
A;Accession: A39355  
A;Molecule type: DNA  
A;Residues: 1-93 <ROS>  
A;Cross-references: GB:M62841  
R;Takada, Y.; Strominger, J.L.; Hemler, M.E.  
Proc. Natl. Acad. Sci. U.S.A. 84, 3239-3243, 1987  
A;Title: The very late antigen family of heterodimers is part of a superfamily of molecu  
A;Reference number: A94151; MUID:87204112; PMID:3033641  
A;Accession: D28018  
A;Molecule type: protein  
A;Residues: 40-50, E', 52-53 <TA2>  
C;Genetics:  
A;Gene: GDB:ITGA4; CD49D  
A;Cross-references: GDB:128032; OMIM:192975  
A;Map position: 2q31-2q32  
C;Superfamily: integrin alpha-4 chain

C;Keywords: cell adhesion; cytoskeleton; duplication; glycoprotein; heterodimer; transmem  
F;1-39/Domain: signal sequence #status predicted <SIG>  
F;40-1038/Product: integrin VLA-4 alpha-4 chain #status predicted <MAT>

Query Match 11.2%; Score 684; DB 2; Length 1038;  
Best Local Similarity 22.9%; Pred. No. 2,6e-42;  
Matches 289; Conservative 181; Mismatches 423; Indels 368; Gaps 54;

QY 9 LFLPLVFLTGLCSFPLNDEHHPLRFPGPPEABFGYSLVQHVGGQRMVLVAP---W-DG 64  
Db 28 LLLCLGVPFG--RPYVNDTESALLYQCPHNTLFGYSLVLSHGANRLLVGAPTANLAN 85  
QY 65 PSQDRGDVYRCVPV-----GAHNA-PCAKGHLGDYQLGNSSHPANVNHGMS 111  
Db 86 ASVINPGAIRCRIGKNKPGOTCEQLQSGNPGPCGKTCLEERD-----NOWLGVT 136  
QY 112 LLETGDD-GGFMACAPLWSRACGSSVSSGICARVDASFOQSLAPTAQRCPTMDVVI 170  
Db 137 LSRQPGENGSIIVTCGRW----- 154  
QY 171 VLDGNSIYPMSEVQTFRLRLVGLFIDPEQIOVLVQYGESPVHWSLGDFTKEEVR 230  
Db 155 ---KNIFYIKNE-----NKLPTGGC-YGVPP-----DLRT----- 180  
QY 231 AAKNLSRREGRETAKTAQIMVACTEGFSQSHGGRPEARLLVVVTDGESHGDELPAALK 290  
Db 181 ----- 187  
QY 291 ACBAGVTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDERFFNVNDEAALTDIVDAL 350  
Db 188 PCYQIVYKFG----- 198  
QY 351 GDRIFGLEGSHAENESSFGLMSQIGFSTRHLKDGILFGMVAGYDMGGSVLWLE---GGH 407  
Db 199 -----ENFAS-----CQAGISSPYTKDLIVMGAPGSSYMTGSLFVYNITNKY 241  
QY 408 RLFPFRMALEDEFPALONH---AAYLYGYSVSSMLLRGRRLL-FLSGAPFRIRGKVIAP 463  
Db 242 KAFLDK-----QNOVKFGSYLGYSVGAGHFRSQHTTEVVGAPQHEQIKAYIF 290  
QY 464 QLKKGDAVRVAQLOGEQISGYFSGELCPDTRDGTDLVLLVAAMPFLGPQN--KETGR 521  
Db 291 SI-DEKELNILEMKKGLGSGYFASVCAVDLNADGFS-DLLVGAPM---QSTIREGR 344  
QY 522 VYVYL-VGQOSSL-TLQGTLPQPEPPQDARFGFANGALPDNLQDGFADVAVGAPLEDGHQ 579  
Db 345 VFVYVINGSGAVNMAMETNLVSGDKYAARFGESIVNLGIDIDNGFEDVAIGAPQEDDLQ 404  
QY 580 ALYLYHGTQSGVPRPAPQRIATAAAMPHALSYFGRSDVGRDLDDGDDLDVAVGA--QGA 637  
Db 405 AIYIYNGRADGISSTFSQRIEGLQISKLSMFGQISGGIDADNNGYVDVAVGAFRSDSA 464  
QY 638 ILLSSRPVHLTLPLEVTPOAIVSVORDCRRRQEAVALTAALCFQVTSR-TPGRWDHQF 696  
Db 465 VLLRTPVIVDASLS-HPESVNRTKFDVCVENGWPSVICIDLTLCFSYKGEVFG----YI 519  
QY 697 YMRFTASLD---EWTAGARAAFDGSG-QRLSPRLRLSVGNVTCEQLH---FHVLDTSYL 750  
Db 520 VLFYNNMSLDVNRKAESPFRFYFSSNGTSDVITGSIQVSSREANC-RTHQAFMRKQVDRIL 578  
QY 751 RPVALTVTAL-----DNTTKPGPVNLGSPSTSI-OKLVPSKCGPDNECVTDL 799  
Db 579 TPIQIEAAVHLGPHVSIKRSSTEEFPLQPTLOQKKEKIMKKTINFARFAHEN-CSADL 637  
QY 800 VLQVND-IRGSRKAPVVRGGRKVLVSTLTENRKENAYNTSLSIIFSRNLHLASLTTPQ 858  
Db 638 QVSAKIGFLPKPHENKTYLAVGSMKTMVSLNAGDDAYETTLNQLPTGLYFIKILDE 697  
QY 859 RESPIKVECAAPSAHARL-CSVGHVPFQ--TGAKVTFLEBFESCSLLSQVFGKLT--- 912  
Db 698 EEKQINCEVTDNSGVQLDCSICGIYVDRSLRIDISFLDV-----SSLSRAHEDLSITV 752  
QY 913 -ASSDSLERNGLQENTAQTSAYIQEPHLLFSSESTLHRYEVP---YGTLPVGFGE 967

Db	753	HATCEEEEMDNLKHSTVVAIPLKYEVKL-----TVHGF-VNPTSFVYGSNDENE--PE	804
Qy	968	-----FKTTLRVQLGCVYVSGLIISALLP-AVAHGGNYFLSLSOVITNNASCIVQNLT	1020
Db	805	TCMVEKMNLTPHVINTGNSMAPVSVIEIMVPNSFSQPTDKLFILVDQITTGEC-----	858
Qy	1021	EPGPPVHPBELQHTNRLNGSNTQCQVVR-----CHLG	1053
Db	859	-----HFENYQRVCALKEQKSAQMLKGIIVRFLSKTRKLLLYCIKADPHCLNLCNFG	911
Qy	1054	QLAKGTEVSVGLLRLVHNEFPFRAPKFSLVST--PEL---GTEEGS--VLQITEASRW	1106
Db	912	KMESGKEAS-----VHIQLEGPSILEMDETSALKPFIIRATGFPFPNPVILNKDENV	965
Qy	1107	SESILEVV--QTRPILSLMILGTSVGGLLMALLVFCCLKLGGF--AHKKTPEBEKRE	1162
Db	966	AHVLEGLGHQRPKRYETIIVISSSLGLGLVLLLTISYVMWKAGFFKPKQYSILQENRR	1025
Qy	1163	E 1163	
Db	1026	D 1026	

Search completed: April 6, 2005, 12:16:59  
Job time : 48.6668 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 172.72 Seconds

(without alignments)  
3459.921 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 6106

Sequence: 1 MELPFVTHLPVFLVLTGLC.....GFFAHKKIPBEKREKLEQ 1167

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6106	100.0	1167	1 ITAG_HUMAN	O75578 homo sapien
2	2429	39.8	1189	1 ITAH_HUMAN	Q9ukx5 homo sapien
3	2418	39.6	1188	1 ITAH_MOUSE	P61622 mus musculus
4	2418	39.6	1188	2 Q7QCQ3	Q7qcq3 mus musculus
5	1894.5	31.0	1180	1 ITAL_RAT	P18614 rattus norv
6	1887	30.9	1151	1 ITAL_HUMAN	P56199 homo sapien
7	1853.5	30.4	1171	2 O42054	O42094 gallus gall
8	1756	28.8	1181	1 ITA2_HUMAN	P17301 homo sapien
9	1754	28.7	1170	1 ITA2_BOVIN	P53710 bos taurus
10	1744	28.6	1178	1 ITA2_MOUSE	Q62469 mus musculus
11	1743.5	28.6	1178	2 Q6P1C7	Q6p1c7 mus musculus
12	1487.5	24.4	823	2 Q8WY18	Q8wy18 homo sapien
13	1473	24.1	823	2 Q8CE84	Q8ce84 mus musculus
14	1273	20.8	288	2 Q8BM12	Q8bm12 mus musculus
15	1136	18.6	1161	1 ITAD_RAT	Q9gye7 rattus norv
16	1131.5	18.5	1162	1 ITAD_HUMAN	Q13349 homo sapien
17	1105	18.1	1170	1 ITAL_HUMAN	P20701 homo sapien
18	1097	18.0	1161	2 Q9WTV4	Q9wtv4 mus musculus
19	1097	18.0	1163	1 ITAL_MOUSE	P24063 mus musculus
20	1089.5	17.8	1160	2 Q9R200	Q9r200 mus musculus
21	1085.5	17.8	1188	2 Q6KAS4	Q6kas4 mus musculus
22	1079	17.7	1152	1 ITAM_HUMAN	P11215 homo sapien
23	1069	17.5	1169	1 ITAM_MOUSE	Q9qxh4 mus musculus
24	1054	17.3	1163	1 ITAL_HUMAN	P20702 homo sapien
25	1053.5	17.3	1165	1 ITAL_BOVIN	P61625 bos taurus
26	1026.5	16.8	1166	2 Q6TYB8	Q6tyb8 bos taurus
27	1022.5	16.7	1153	1 ITAM_MOUSE	P05555 mus musculus
28	1000.5	16.4	1151	2 Q9J130	Q9j130 rattus norv
29	997	16.3	780	2 Q06271	Q06271 xenopus lae
30	984	16.1	1086	2 Q96HB1	Q96hb1 homo sapien
31	980.5	16.1	1196	2 Q98TF1	Q98tf1 cyprinus ca

32	953	15.6	1167	1 ITAE_MOUSE	Q60677 mus musculus
33	953	15.6	1167	2 O88340	O88340 rattus norv
34	947.5	15.5	1187	2 Q98TF0	Q98tf0 cyprinus ca
35	946	15.5	1167	2 O88341	O88341 rattus norv
36	932.5	15.3	1179	1 ITAE_HUMAN	P38570 homo sapien
37	891.5	14.6	1038	2 Q8BS01	Q8bs01 mus musculus
38	883.5	14.5	920	2 Q28984	Q28984 sus scrofa
39	883	14.5	895	2 Q9WJF8	Q9wuf8 mus ep. itg
40	863	14.1	927	2 Q8HZV0	Q8hzv0 bos taurus
41	797	13.1	1332	2 Q9BP08	Q9bpq8 halocynthia
42	787.5	12.9	1160	2 Q8MKF4	Q8mkf4 felis silve
43	778	12.7	1036	2 Q91YD5	Q91yd5 mus musculus
44	738	12.1	1035	1 ITA9_HUMAN	Q13797 homo sapien
45	732.5	12.0	1054	2 Q9U6S1	Q9u6s1 strongyloce

#### ALIGNMENTS

#### RESULT 1

ID	ITAG_HUMAN	STANDARD;	PRT;	1167 AA.
AC	O75578; Q9UH28;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	05-JUL-2004 (Rel. 44, Last annotation update)			
DE	Integrin alpha-10 precursor.			
GN	Name=ITGA10;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Articular chondrocytes;			
RX	MEDLINE=98352078; PubMed=9685391; DOI=10.1074/jbc.273.32.20383;			
RA	Camper L., Hellman U., Lundgren-Akerlund E.,			
RT	"Isolation, cloning, and sequence analysis of the integrin subunit			
RT	alpha10, a beta1-associated collagen binding integrin expressed on			
RT	chondrocytes.";			
RL	J. Biol. Chem. 273:20383-20389(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Endothelial cells, and Heart;			
RX	MEDLINE=20169197; PubMed=10702680;			
RA	Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,			
RA	Wang S.-X., Langley R., Krissansen G.W.;			
RT	"The integrin alpha10 subunit: expression pattern, partial gene			
RT	structure, and chromosomal localization.";			
RL	Cytogenet. Cell Genet. 87:238-244(1999).			
CC	-!- FUNCTION: Integrin alpha-10/beta-1 is a receptor for collagen.			
CC	-!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-10			
CC	associates with beta-1.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-!- TISSUE SPECIFICITY: Widely expressed with highest expression in			
CC	muscle and heart. Found in articular cartilage.			
CC	-!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins			
CC	with I-domains do not undergo protease cleavage.			
CC	-!- SIMILARITY: Belongs to the integrin alpha chain family.			
CC	-!- SIMILARITY: Contains 7 FG-GAP repeats.			
CC	-!- SIMILARITY: Contains 1 VWFA domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL; AF074015; AAC31952.1; -			
DR	EMBL; AF112345; AAF21944.1; -			
DR	EMBL; AF172723; AAF61638.1; -			

```

DR HSP; P18614; IMHP.
DR Genew; HGNC:6135; ITGA10.
DR MIM; 604042; -.
DR GO; GO:0008105; C:integrin complex; TAS.
DR GO; GO:0005518; P:collagen binding; TAS.
DR GO; GO:0007160; P:cell-matrix adhesion; TAS.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; VWF_A.
DR Pfam; PF00092; VWA_1.
DR PRINTS; PRO1185; INTEGRIN.
DR PRINTS; PRO0453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA_1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
DR PROSITE; PS02034; VWA; 1.
DR Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;
KW Repeat; Signal; Transmembrane.
FT SIGNAL 1 22 Potential.
FT CHAIN 23 1167 Integrin alpha-10.
FT DOMAIN 23 1122 Extracellular (Potential).
FT TRANSMEM 1123 1145 Potential.
FT DOMAIN 1146 1167 Cytoplasmic (Potential).
FT REPEAT 38 97 FG-GAP 1.
FT REPEAT 98 ? FG-GAP 2.
FT DOMAIN 167 350 VWA.
FT REPEAT 365 427 FG-GAP 3.
FT REPEAT 428 482 FG-GAP 4.
FT REPEAT 483 545 FG-GAP 5.
FT REPEAT 546 605 FG-GAP 6.
FT REPEAT 608 660 FG-GAP 7.
FT DOMAIN 1134 1140 Poly-Leu.
FT CA_BIND 494 502 Potential.
FT CA_BIND 558 566 Potential.
FT DISULFID 76 86 By similarity.
FT DISULFID 666 675 By similarity.
FT DISULFID 681 736 By similarity.
FT DISULFID 789 795 By similarity.
FT CARBOHYD 98 98 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 234 234 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 336 336 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 364 364 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 733 733 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 763 763 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 839 839 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 921 921 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 1011 1011 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 1018 1018 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 1039 1039 N-linked (GlcNAc... ) (Potential).
FT CONFLICT 844 844 I -> L (in Ref. 2).
FT CONFLICT 909 909 G -> V (in Ref. 2).
FT CONFLICT 926 926 E -> D (in Ref. 2).
SQ SEQUENCE 1167 AA; 127573 MW; AE7D3A1C25C1AE0 CRC64;

Query Match
Best Local Similarity 100.0%; Score 6106; DB 1; Length 1167;
Matches 1167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MELPFVTHLFLPLVLTGLCSFFNLDEHHPLFPQPEAEFGYSVLQHVGGQRWMLVGA 60
Db 1 MELPFVTHLFLPLVLTGLCSFFNLDEHHPLFPQPEAEFGYSVLQHVGGQRWMLVGA 60

Qy 61 PWDPSGDRRDVRCVPVGGAHNAPCAKXGHLGDLQGNSSHPAVNMHLGMSLLTDDGG 120
Db 61 PWDPSGDRRDVRCVPVGGAHNAPCAKXGHLGDLQGNSSHPAVNMHLGMSLLTDDGG 120

Qy 121 FMACAPLWSRACGVSFSSGICARVDASFPQGSILAPTAQRCPYMDVIVLDGNSIYP 180
Db 121 FMACAPLWSRACGVSFSSGICARVDASFPQGSILAPTAQRCPYMDVIVLDGNSIYP 180

Qy 191 WSEVQTFRLRLVGKLFIDPEQIQVGLVQGESPVHWSLGDFTKBEVVRAAKNLSRREG 240

```

```

Db 181 WSEVQTFRLRLVGKLFIDPEQIQVGLVQGESPVHWSLGDFTKBEVVRAAKNLSRREG 240
Qy 241 RETKTAQIMVACTEGFSQSHGGRPAARLLVVVTGESHGDBELPAALKACAGRVTRY 300
Db 241 RETKTAQIMVACTEGFSQSHGGRPAARLLVVVTGESHGDBELPAALKACAGRVTRY 300
Qy 301 GIAVLGHYLRQRDPSSFLREIRTIASDPDERFFFNVTDEAALTDIVDALGDRIFGLEGS 360
Db 301 GIAVLGHYLRQRDPSSFLREIRTIASDPDERFFFNVTDEAALTDIVDALGDRIFGLEGS 360
Qy 361 HAENESSFGLEMSQIGFSTHRLKDGILFGMVGYDNGGSLVLMLEGGHRLPPRMALEDEF 420
Db 361 HAENESSFGLEMSQIGFSTHRLKDGILFGMVGYDNGGSLVLMLEGGHRLPPRMALEDEF 420
Qy 421 PPALQNHAAVLYGYSVSMMLRGRRFLFSGAPFRHGRKVIAPQLKKDGAVRVAQSLQGE 480
Db 421 PPALQNHAAVLYGYSVSMMLRGRRFLFSGAPFRHGRKVIAPQLKKDGAVRVAQSLQGE 480
Qy 481 QIGSYFGSELCPDTRDGTGTVLLVAAMPFLGPQNKETGRVYVYLVGQOSSLTLQGTQL 540
Db 481 QIGSYFGSELCPDTRDGTGTVLLVAAMPFLGPQNKETGRVYVYLVGQOSSLTLQGTQL 540
Qy 541 PEPPQDARFGFAMGALPDNLNODGFADVAVGAPLEDGHQGALYLYHGTQSVRPHAPQRIA 600
Db 541 PEPPQDARFGFAMGALPDNLNODGFADVAVGAPLEDGHQGALYLYHGTQSVRPHAPQRIA 600
Qy 601 AASMPHALSYFGRSVDGRDLDDGDDLDVDAVAGAGAAILLSSRPVHLTSPLEVTPOAIS 660
Db 601 AASMPHALSYFGRSVDGRDLDDGDDLDVDAVAGAGAAILLSSRPVHLTSPLEVTPOAIS 660
Qy 661 VVQDCRRRQGEAVCLTAALCFQVTSRTPGRWDPHFYMRFTASLDEWTAGARAAFDGSGQ 720
Db 661 VVQDCRRRQGEAVCLTAALCFQVTSRTPGRWDPHFYMRFTASLDEWTAGARAAFDGSGQ 720
Qy 721 RLSPRRLRLSGVNTCEQLHFHVLDTSDYLRLPVALTVPALDNTTKPGPVINEGSPISIQ 780
Db 721 RLSPRRLRLSGVNTCEQLHFHVLDTSDYLRLPVALTVPALDNTTKPGPVINEGSPISIQ 780
Qy 781 KLVPPSKDCGPDNECVTDLVQNMIDIRGSRKAPFVVRGRRKVLVSTTLLENKENAYNT 840
Db 781 KLVPPSKDCGPDNECVTDLVQNMIDIRGSRKAPFVVRGRRKVLVSTTLLENKENAYNT 840
Qy 841 SLSIIFSRNLHLASLTLPQRESPIKVECAAPSAHARLCSVGHVPVQTKGAKVTFLEPFESC 900
Db 841 SLSIIFSRNLHLASLTLPQRESPIKVECAAPSAHARLCSVGHVPVQTKGAKVTFLEPFESC 900
Qy 901 SLLSQVFGKLTASSDSLENGTLQENTATQTSAYIQEPHLLFSSESTLHRYEVHPYGTL 960
Db 901 SLLSQVFGKLTASSDSLENGTLQENTATQTSAYIQEPHLLFSSESTLHRYEVHPYGTL 960
Qy 961 PVGPGPFKTLRVQNLGCVVWSGLIISALLPAVAGGNYPFLSLSQVITNNASCIVQNL 1020
Db 961 PVGPGPFKTLRVQNLGCVVWSGLIISALLPAVAGGNYPFLSLSQVITNNASCIVQNL 1020
Qy 1021 EPPGPPVHPELOHTNRLNGSNTCCQVVRCHLQAGTSEVSGLLRLVHNEFFRRAKFK 1080
Db 1021 EPPGPPVHPELOHTNRLNGSNTCCQVVRCHLQAGTSEVSGLLRLVHNEFFRRAKFK 1080
Qy 1081 SLTVVSTFELGTREGSVLQTEASRWSESLELVQVTRPILISWILIGSVLGGILLALL 1140
Db 1081 SLTVVSTFELGTREGSVLQTEASRWSESLELVQVTRPILISWILIGSVLGGILLALL 1140
Qy 1141 VFCLMKLGFFAHKIPPEEKREKLEQ 1167
Db 1141 VFCLMKLGFFAHKIPPEEKREKLEQ 1167

RESULT 2
ITAH HUMAN
ID ITAH_HUMAN STANDARD; PRT; 1189 AA.
AC Q9UUK5; Q9UUK1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```



05-JUL-2004 (Rel. 44, Last annotation update)  
 Integrin alpha-11 precursor.  
 Names:ITG11;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;  
 [1]  
 SEQUENCE FROM N.A.  
 RC TISSUE=Fetal heart, and Osteoblast;  
 RX MEDLINE=99417678; PubMed=10486209; DOI=10.1006/geno.1999.5909;  
 RA Lehnert K., Ni J., Leung E., Gough S.M., Weaver A., Yao W.P., Liu D.,  
 Wang S.-X., Morris C.M., Kriisaenen G.W.;  
 RT "Cloning, sequence analysis, and chromosomal localization of the novel  
 human integrin alpha11 subunit (ITG11).";  
 RL Genomics 60:179-187(1999).  
 [2]  
 SEQUENCE FROM N.A.  
 RC TISSUE=Fetal muscle, and Uterus;  
 RX MEDLINE=99395147; PubMed=10464311; DOI=10.1074/jbc.274.36.25735;  
 RA Velling T., Kusche-Gullberg M., Sejersen T., Gullberg D.;  
 RT "cDNA Cloning and Chromosomal Localization of Human alpha(11)  
 Integrin. A collagen-binding, i domain-containing, beta(1)-associated  
 integrin alpha-chain present in muscle tissues.";  
 RL J. Biol. Chem. 274:25735-25742(1999).  
 [3]  
 SEQUENCE OF 954-1188 FROM N.A.  
 RC TISSUE=Fibroblast;  
 RA Andreu N., Estivill X., Escarceller M., Sumoy L.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Integrin alpha-11/beta-1 is a receptor for collagen.  
 CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-11  
 CC associates with beta-1.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: According Ref.2 highest levels in uterus and  
 CC heart, intermediate levels in skeletal muscle and intermediate to  
 CC low levels in pancreas, kidney and placenta. According to Ref.1  
 CC also found in brain, colon, lung, small intestine, stomach,  
 CC testis, salivary glands, thyroid glands and prostate. Very low  
 CC levels in peripheral blood lymphocytes, fetal brain and fetal  
 CC liver.  
 CC -1- DEVELOPMENTAL STAGE: Strongly up-regulated in differentiating  
 CC fetal muscle cells (in vitro).  
 CC -1- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins  
 CC with I-domains do not undergo protease cleavage.  
 CC -1- SIMILARITY: Belongs to the integrin alpha chain family.  
 CC -1- SIMILARITY: Contains 7 FG-GAP repeats.  
 CC -1- SIMILARITY: Contains 1 VWFA domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
 CC  
 CC EMBL; AF109681; AA01258.1; -  
 CC EMBL; AF137378; AA051919.2; -  
 CC EMBL; AL359064; CAB94392.1; -  
 CC HSP; P18614; IMHP.  
 CC Genew; HGNC:6136; ITG11.  
 CC MIM; 604789; -  
 CC GO; GO:0008305; C:integrin complex; TAS.  
 CC GO; GO:0005518; F:collagen binding; TAS.  
 CC GO; GO:0007160; P:cell-matrix adhesion; TAS.  
 CC GO; GO:0007517; P:muscle development; TAS.  
 CC InterPro; IPR00413; Integrin\_alpha.  
 CC InterPro; IPR002035; FG-GAP; 3-  
 CC Pfam; PF01839; FG-GAP; 3-  
 CC Pfam; PF00092; VWFA; 1-  
 CC PRINTS; PRO1185; INTEGRINA.  
 CC PRINTS; PRO0453; VWFADOMAIN.

DR SMART; SM00191; Int\_alpha; 5.  
 DR SMART; SM00327; VWFA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; FALSE\_NEG.  
 DR PROSITE; PS0234; VWFA; 1.  
 KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium;  
 KW Polymorphism; Receptor; Repeat; Signal; Transmembrane.  
 FT SIGNAL 1 22 Potential  
 FT CHAIN 23 1189 Integrin alpha-11.  
 FT DOMAIN 23 1142 Extracellular (Potential).  
 FT TRANSMEM 1143 1165 Potential.  
 FT DOMAIN 1166 1189 Cytoplasmic (Potential).  
 FT REPEAT 38 94 FG-GAP 1.  
 FT REPEAT 102 163 FG-GAP 2.  
 FT DOMAIN 164 345 VWFA.  
 FT REPEAT 359 420 FG-GAP 3.  
 FT REPEAT 422 475 FG-GAP 4.  
 FT REPEAT 477 537 FG-GAP 5.  
 FT REPEAT 539 598 FG-GAP 6.  
 FT REPEAT 601 653 FG-GAP 7.  
 FT DOMAIN 1154 1162 Poly-Leu.  
 FT DOMAIN 1174 1177 Poly-Arg.  
 FT CA\_BIND 488 496 Potential.  
 FT CA\_BIND 551 559 Potential.  
 FT CA\_BIND 613 621 Potential.  
 FT DISULFID 76 83 By similarity.  
 FT DISULFID 121 139 Potential.  
 FT DISULFID 129 159 Potential.  
 FT DISULFID 659 668 By similarity.  
 FT DISULFID 674 729 By similarity.  
 FT DISULFID 781 787 By similarity.  
 FT DISULFID 881 893 By similarity.  
 FT CARBOHYD 82 82 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 95 95 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 291 291 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 331 331 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 358 358 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 449 449 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 462 462 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 528 528 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 642 642 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 694 694 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 857 857 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 894 894 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 973 973 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1032 1032 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1040 1040 N-linked (GlcNAc... ) (Potential).  
 FT VARIANT 433 433 V -> M.  
 FT VARIANT 524 524 /FTid=VAR\_009889.  
 FT VARIANT 524 524 R -> L.  
 FT VARIANT 972 972 /FTid=VAR\_009890.  
 FT VARIANT 972 972 L -> P.  
 FT VARIANT 1003 1003 /FTid=VAR\_009891.  
 FT VARIANT 1003 1003 I -> M.  
 FT VARIANT 1030 1030 /FTid=VAR\_009892.  
 FT VARIANT 1030 1030 Missing.  
 FT VARIANT 1094 1094 /FTid=VAR\_009893.  
 FT VARIANT 1094 1094 L -> V.  
 FT VARIANT 1094 1094 /FTid=VAR\_009894.  
 SQ SEQUENCE 1189 AA; 133609 MW; 60303C08AA4CD52 CRC64;  
 Query Match 39.8%; Score 2429; DB 1; Length 1189;  
 Best Local Similarity 43.1%; Pred. No. 3.5e-164;  
 Matches 516; Conservative 213; Mismatches 422; Indels 46; Gaps 16;  
 QY 1 MELPFTVTHLPLVFLTGLCSFNLDEHHPRLPFPPEAEFGYSVLQHVGGQRMVLVGA 60  
 DB 1 MDLPGRGLVVAWLSLWFGFTDTTFNMDRKFRVIFGSTRFAFGYTVQQHDSGNKWL VVGA 60  
 QY 61 PWGPGSGDRRGDYYRCVPGVGGAHNAPCAKGHLDYQLGNSSHPAVNMHLGSLLETG DGG 120  
 DB 61 PLETNQYQKTGDVYKCPV---IHGNC TKLNLGRVTLNSVSKDNMRGLSLATNP KDNS 117  
 QY 121 FMACAPLWSRACGSSGVFSSSGICARVDASFQPGSLAPTAQRCPYMDVIVLDG SNIYP 180

Db 118 FLACPLWSEHCSSYTTGMCSSVNGNFRFSKTVAPALQRCQTYMDIVVLGDSNIYP 177  
 QY 181 WSEVQTLRLRLVGLKFLDPEQIOVLGVQYESPVHWSLGDFTKKEVRAAKNLSREG 240  
 Db 178 WVEVQHFLIILKKFYIGPGQIQGVVQYQVEDVHFEHLNDYRSVKDQVVAASHIBORGG 237  
 QY 241 RETKTAQIAIMVACTGFSQSGHGRPEAARLLVVVVTDGESHGDBELPAALKACAGRVTRY 300  
 Db 238 TETRATGIFSFARSEAFQK--GGRKGAKVMIVITDGESHSDPLEKVIQOQSERDNTVRY 295  
 QY 301 GIAVLGHLRQRDPSSFLREIRTIASDPDRPFNFVTDDBAALTDIVDALGDRIPGLEGS 360  
 Db 296 AVAVLGYNRRGINPETFLNEIKYIASDPDKHFNVTDEAALKDIVDALGDRIFSLGEGT 355  
 QY 361 HAENESSFGLMSQIGFSTRHLKDKGILFGWVGAVDMGSLVLMLEGGHRLPPRMALBDEF 420  
 Db 356 N-KNETSFGLEMSGTGFSSHVVEDVLGAVGAVDMGAVLKTSAGKVIPLRESYLKEF 414  
 QY 421 PPALQNHAAVLGYSSVSMLLRGRRFLSLGAPRPHRGKVIATFOLKKDGAVRVAQSLQGE 480  
 Db 415 PEELKXHGAVLGYTTSVSSRQGRVYVAGAPRPHHTGKVLFTMHNRSLSLTHQAMRGQ 474  
 QY 481 QIGSYFSGELCPDLTDGDTDLVLLVAPMLPGQNKETGRVYVVLVGGQSLTLTLOQTLLQ 540  
 Db 475 QIGSYFSGEITSVDIDGQVTDVLLVGAPMYFN--EGRERGKVVYVEL--RQNRFDVYNGTLK 532  
 QY 541 PEPP-QDARGFGAMGALPDNLQDGFADVAVGAPLEDHOGALVLYHGTQSGVRPHRQRI 599  
 Db 533 DSHSYQNARFGSSIASVRDLNQSDINDVVGAPLEDNHAGAIYIFHGFSGSLKTPQRI 592  
 QY 600 AAASMPHALSYFGRSVDRGLDGLDDLDVAVAGQAAILLSRPIVHLTPSLSEVTPQAI 659  
 Db 593 TASELATGLQYFGCSINGQLDNEGLDLAVGALGNVILMSRPVVOINASHLFESKI 652  
 QY 660 SVVORDCRRGQENAVCLTAALCFQVTSRTQGRMDHQPMFTASLDWTGARAAPFGSG 719  
 Db 653 NIFHRDKCRSGRDATCLAAFLCFTPIFLAPHFQTTVIGIRYNATMDERRVTPRAHLDEGG 712  
 QY 720 QRLSPRLRLSVGNVTCEQLHFEHLVDTSDYLRLPVALTFTALDNTTKPGVILNBSPTS 779  
 Db 713 DRFTNRAVLSSGQELCERINFHVLDYADVVKPFTSVSEYSLDPDH--GPMLDGQWTTLL 771  
 QY 780 QKLVPPSKDCPDNECVTDVLQVNMDI-----RGSRK-----APFVV 817  
 Db 772 RVSVFPWNGCNEDEHCVPLDLVLDARSDLPTAMEYQQRVLRKPAQDCSAYTSLSPDTTVFII 831  
 QY 818 RGRKRVLVSTTLNRKENAYNTSLSIIFGRNLHLASLTQORSPIKVECAAPS--AHAR 875  
 Db 832 ESTQRVAVEATLENRGENAYSTVNTSQSANLQFASLIQKEDSDGSEICVNBERRLQKQ 891  
 QY 876 LCSYGHVPVFTGAKVTFLLPEFSCSLLSQVFGKLTASSDSLERNCTLOENTATQTSAYI 935  
 Db 892 VCNVSYFFRAKAVARLDFEFESKIFLHLELTAAGSDSNRSDTKSDNAPLRFHL 951  
 QY 936 QYEPHLLFSSESTLHRYEVHPYGTLP--VGPGEFTKTLRVQLNGCVVSGLIISALLPA 993  
 Db 952 KYEADVLFTRSSLSHVEKLNLSLERYDGIQPPFSCIPIQNLGLPFIHGMKTIPI 1011  
 QY 994 VAHGGNVFLSQQVITN--NASC-IVQNTTEPPGPPVHPEELQHTNRLNGSNTQCCOVR 1049  
 Db 1012 ATRSGNRLKLRLDPLTDEVANTSCNIWGNSTEVPTPE--EDLRRAPQLNHSNDVVSIN 1070  
 QY 1050 CHLQOLAKGTEVSGLLRLVHNEFFRAKPSLTVVSTFELGTBEGSVLQLTASRWSSES 1109  
 Db 1071 CNI--RLVNPQINPHLLGNLWRLSKALKYKSMKIMVNAALQROFHPFFTFREEDPSRQI 1129  
 QY 1110 LLEVQTRPILISWLIGVLGGLLLALLVFCMLKGLFFAHKKIPEEKKREKLE 1166  
 Db 1130 VFEISKQEDMQVPIWIIVGSTLGLLLALLVLLALMLKLGFRSAR----RRREPLD 1182

ITAH MOUSE  
 ID ITAH MOUSE STANDARD; PRT; 1188 AA.  
 AC P61622;  
 DT 05-JUL-2004 (Rel. 44, Created)  
 DT 05-JUL-2004 (Rel. 44, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Integrin alpha-11 precursor.  
 GN Name=Itgall;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh P.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnarch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: Integrin alpha-11/beta-1 is a receptor for collagen (By  
 CC similarity).  
 CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-11  
 CC associates with beta-1 (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins  
 CC with I-domains do not undergo protease cleavage.  
 CC -!- SIMILARITY: Belongs to the integrin alpha chain family.  
 CC -!- SIMILARITY: Contains 7 FG-GAP repeats.  
 CC -!- SIMILARITY: Contains 1 VWFA domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; BC058716; AAH58716.1; -;  
 CC MGD; MGI:2442114; Itgall.  
 CC PROSITE; PS00242; INTEGRIN\_ALPHA; FALSE\_NEG.  
 CC PROSITE; PS0234; VWFA; 1.  
 CC Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;  
 CC Repeat; Signal; Transmembrane.  
 CC SIGNAL 1 22 Potential.  
 CC CHAIN 23 1188 Integrin alpha-11.  
 CC DOMAIN 23 1141 Extracellular (Potential).  
 CC TRANSMEM 1142 1164 Potential.  
 CC DOMAIN 1165 1188 Cytoplasmic (Potential).  
 CC REPEAT 38 94 FG-GAP 1.  
 CC REPEAT 102 163 FG-GAP 2.  
 CC REPEAT 164 345 VWFA.  
 CC DOMAIN 359 420 FG-GAP 3.  
 CC REPEAT 422 475 FG-GAP 4.  
 CC REPEAT 477 537 FG-GAP 5.  
 CC REPEAT 539 598 FG-GAP 6.  
 CC

FT	REPEAT	601	653	FG-GAP 7.	QY	617	GRLDGGDLVAVGAAGAAIILSSRPVHLTPSLEVTPQAISVVDORCHRRGOEAVCL	676
FT	CA_BIND	488	496	Potential.	Db	610	QGLDNEGLVLAVALGALVAVVLPVQINSLHFEFSKINIFHKDKRGRDACL	669
FT	CA_BIND	551	559	Potential.	QY	677	TAALCFQVTSRTGRWDHQFYMRFRTASLDEWTAAGARAAFDGSGQRLSPRLRLSVGNVTC	736
FT	CA_BIND	613	621	Potential.	Db	670	AAFLCFPIPIFLAHPFQTATVIGIRYNATMDERRYMPRAHLDGGQDFTNRAVLSSGQSHC	729
FT	DISULFID	76	83	Potential.	QY	737	EQHFFHVLDSYLRPVVALVTTFALDNTTTPGPGVNLGSPSTSIQKLVFPFKDCQPDNECV	796
FT	DISULFID	121	139	Potential.	Db	730	QRINFHVLDTADYKVPVAFSVYSLEDPDN-GPMLDNGWPTTLRVSVFPMWNCNEDEHCV	788
FT	DISULFID	129	159	Potential.	QY	797	TDLVQLVQNMDIRSK-----APFVVRGGRRKVLVSTLTLENRK	834
FT	DISULFID	674	729	By similarity.	Db	789	PDVLVDARSULPTAMEYQCVLRPAQDCSSYTLSTFTTFTIESTRRRVAVEATLENRG	848
FT	DISULFID	781	787	By similarity.	QY	835	ENAYNTSLSIIFSRNLHLASLTPORESPIKVECAAPS--AHARLCSVGHVPVFTQCAKVTFF	892
FT	DISULFID	881	893	By similarity.	Db	849	ENAYSAVINISQSENLFQASLIQKDDSDNSECNEERRLHKVCNVSYPPFRAKAKVAF	908
FT	CARBOHYD	82	95	N-linked (GlcNAc. .)	QY	893	LLPFESCSSLQVFOKLTASSDSLRNGTLQNTAQTSAIYQYEPHLLFSSSTLHRY	952
FT	CARBOHYD	95	95	N-linked (GlcNAc. .)	Db	909	RLDFEFSKSVFLHHLQIHLGAGSDSHEQDSTADNTALLRFLHLYEADVLFTRSSLSHF	968
FT	CARBOHYD	291	331	N-linked (GlcNAc. .)	QY	953	EVHPYGTLPV--GPGPEFKTTLRVQNLGCVVSGLIISALLPAVAHGSNYFLSLSQVITN	1010
FT	CARBOHYD	331	331	N-linked (GlcNAc. .)	Db	969	EVKANSLESYDGIIGPPFNCVFKVQNLGFFPIHGVMKMTVPIATRGNNLLMLURDFTD	1028
FT	CARBOHYD	358	358	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	449	449	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	462	462	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	528	528	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	528	528	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	642	642	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	694	694	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	857	857	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	894	894	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	973	973	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	1031	1031	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	QY	1068	LVNNEFFRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLELVVQTRPILSLMILI	1127
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1087	NLMTSLTKALKYRSKITVNAALQRPSPFIREFDPQRVTEISKQEDQWQVPIWIIIV	1146
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1128	GSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK	1160
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1147	GSTLGGLLALLVLAWLKLGFFAKRKREPGGLGPIPKELK	1188
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	QY	1011	--NASC-IVONLTPEPPVHPBELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLR	1067
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	Db	1029	QGNWSCNINWNSYESTPTPE-EDLSHAPQNHNSNDVSIICNL-RLASQETSFYLVG	1086
FT	CARBOHY							

```

DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00092; VWA; 1.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS0234; VWA; 1.
KW Cell adhesion; Integrin; Transmembrane.
SQ SEQUENCE 1188 AA; 133067 MW; 69B2BF20D75E4271 CRC64;

Query Match 39.68; Score 2418; DB 2; Length 1188;
Best Local Similarity 43.38; Pred. No. 2.2e-163;
Matches 512; Conservative 207; Mismatches 413; Indels 50; Gaps 15;

QY 18 GLCSFNLDBHRLPFGPPEAEFGYSVLQGVGGQRMVLGAPWDGSPGDRGDVYRCP 77
DB 18 GFTDTFNMDRNPRVIAAGPSAAPFGYTVQQHDISGKKLVVVGAPMETNGHKQTKDGVYKCP 77
QY 78 VGGAHNAPCAKHGLDYLQGNSSHPAVNMHLGHSLLSTDDGGGFMACAPLWSRACGSSVF 137
DB 78 VTQGN---CTKJNLGRVYTLGNVSRKDNMRLGLSLATNPKNDSNPLACPLWSHSECGSSYY 134
QY 138 SSGICARVDASFOFGSLAPTAQRCPTYMDVIVLDGNSIYPWSEVQTFLLRLRLVGKLF 197
DB 135 TTGMSRVNFRFSKIVAPALQRCQTYMDIVIVLDGNSIYPWSEVQTFLLRLRLVGKLF 194
QY 198 DPBQIQVLGYGESPVHWSLGDFTKEEVRAAKNLSRREGREKTXAQAIMVACTEGF 257
DB 195 GPQIQVIGVOYGEDA VHEPLNDYRSVKDVEAASHIEQGGTETRTAFGIEPAREAP 254
QY 258 SQSHGGRPEARLLVVTDGSHDGEELPAALKACEAGRVTYRGIAVLGHYLRQRDPSS 317
DB 255 QK--GGRKGAKMIVITDGHSDSPLEKVIQSEKXNDVTRYAVAVLYGYNRRGINPET 312
QY 318 FLREIRTIASDPPEPFNTVDEAALTDIVDALGDRIFGLEGSHAENESSFGLEMSOIGF 377
DB 313 FLNEIKIYASDPDKHFNFTVDEAALTDIVDALGDRIFGLEGN-KNETSFGLEMSGTGF 371
QY 378 STHRLKDGILFGMVGAYDWMGSLVLEGGHRLFPFRMALEDEFPALQNAHAALVGYSVS 437
DB 372 SSWVEDGILLGAVGAYDMNGAVLKETSAGKVIPIHRESYLKEPPEELKNHVAIVLYTVTS 431
QY 438 MLRGGRRLFLSGAPRRHKGKVIJAFQKKGAVRVAQSIQGGQIGSYFGSELCPLDTR 497
DB 432 VSSRQGRVTVAGAPRPHNTGKVLFSMHNRSLSITHQALRGEQIGSYFGSEITSVDVD 491
QY 498 DGTDTDLVLAAPFLGQNKTRGVYVLYVQGSLLTLQGTLOPEPP-QDARFGFANGAL 556
DB 492 DRTVDULLVGAPMYS-EGERGKVIYVNL-RQNRVYVNTGLKDSHYQNAFPGCSIASV 549
QY 557 PDLNQGFADVAVGAPLEDHQGALYLYHGTQSGVRPHPAQRIAAASMPHALSYFGRSVD 616
DB 550 QDLNQDSYNDVVVGAPLEDSHRGAIYIFHGFTNLKPKMQRITASELAPQLQHFGCSIH 609
QY 617 GRLLDQDLDLVAVGAGQAALLSSRPVHLTPSLVTPQATISVVORDCRRRQBAVCL 676
DB 610 GOLDLNEDGVLAVGALGNAVLMARPVVQINASLHFEPSKINIFPKDCKRNGRDATCL 669
QY 677 TAALCFQVTSKTPGRWDHQFMRPTASLDSEWAGARAAAFDGGSGORLSPRLRLSVGNVTC 736
DB 670 AAFLCFTPIFLAHFQATVIGIRYNALMDERRYPRAHLDEGGDQFTNRAVLSSGQEH 729
QY 737 EQLHFHVLDTSDYLRPVALTATFALDNTKPGPVINEGSPSTISQKLVFPGKDCQPDNECV 796
DB 730 QRINFHVLDTADYVKVPVAFSVSEYSLSDPDN-GPMLDNGWPTTLRVSVFPWNGCNEDEHCV 788
QY 797 TDLVLQVNMIRGRK-----APFVVGGRKRVLVSTLTENRK 834
DB 789 PDLVLDARSDLPTAMEYCORVLRPAQDCSSYTLSPDTTFTIESTRRRAVEATLENRG 848
QY 835 ENAYNTSLSTIFSNNLHSLTTPQRESPIKVECAAPS--AHARLCVSGHPVFOFGAKVTF 892
DB 849 ENAYSAVLNTSQSENLOFASLIQKDDSDNSIECVNEERLHKVCNVSYPFFRAKAVAF 908
QY 893 LLEFEFCSLLSQVFGKLTASSDSLRNGTLQENTAGTSAYIQYEPLLFSSESTLHRY 952

```

```

RESULT 5
ID _ITALI_RAT STANDARD; PRT; 1180 AA.
AC P18614;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Integrin alpha-1 precursor (Laminin and collagen receptor) (VLA-1)
DE (CD49a).
GN Name=Itgal;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=90338125; PubMed=2380249; DOI=10.1083/jcb.111.2.709;
RA Ignatius M.J., Lange T.H., Houde M., Tawil J.W., Barton A., Esch F.,
RA Carbonetto S., Reichardt L.F.;
RT "Molecular cloning of the rat integrin alpha 1-subunit: a receptor for
RT laminin and collagen."
RL J. Cell Biol. 111:709-720(1990).
RN [2]
X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 151-364.
RX MEDLINE=99313197; PubMed=10386626; DOI=10.1016/S0014-5793(99)00666-3;
RA Nolte M., Pepinsky R.B., Vanyaminov S.Y., Kotliansky V.,
RA Gotwals P.J., Karpusas M.;
RT "Crystal structure of the alpha1beta1 integrin I-domain: insights into
RT integrin I-domain function."
RL FEBS Lett. 452:379-385(1999).
CC -I- FUNCTION: Integrin alpha-1/beta-1 is a receptor for laminin and
CC collagen. It recognizes the proline-hydroxylated sequence G-P-P-G-
CC E-R in collagen.
CC -I- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-1
CC associates with beta-1.
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC -I- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -I- SIMILARITY: Belongs to the integrin alpha chain family.
CC -I- SIMILARITY: Contains 7 FG-GAP repeats.
CC -I- SIMILARITY: Contains 1 VWFA domain.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; X52140; CAA36384.1; -.
CC PIR; A35854; A35854.
CC PDB; 1CK4; X-ray; A/B=167-364.

```

DR PDB; 1MHP; X-ray; A/B=169-360.  
 DR RGD; 2923; Itgal.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 3.  
 DR Pfam; PF00357; Integrin\_alpha; 1.  
 DR Pfam; PF00092; VWA; 1.  
 DR PRINTS; PRO1185; INTEGRIN.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 5.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS0234; VWF; 1.  
 DR 3D-structure; Calcium; Cell adhesion; Glycoprotein; Integrin;  
 KW Magnesium; Receptor; Repeat; Signal; Transmembrane.  
 FT SIGNAL 1 28  
 FT CHAIN 29 1180 Integrin alpha-1.  
 FT DOMAIN 29 1142 Extracellular (Potential).  
 FT TRANSMEM 1143 1165 Potential.  
 FT DOMAIN 1166 1180 Cytoplasmic (Potential).  
 FT REPEAT 104 103 FG-GAP 1.  
 FT REPEAT 144 ? FG-GAP 2.  
 FT REPEAT 175 388 VWF.  
 FT REPEAT 377 432 FG-GAP 3.  
 FT REPEAT 433 484 FG-GAP 4.  
 FT REPEAT 485 565 FG-GAP 5.  
 FT REPEAT 567 626 FG-GAP 6.  
 FT REPEAT 629 681 FG-GAP 7.  
 FT CA\_BIND 497 505 Potential.  
 FT CA\_BIND 579 587 Potential.  
 FT CA\_BIND 641 649 Potential.  
 FT SITE 1168 1172 GFFRR motif.  
 FT DISULFID 82 92 By similarity.  
 FT DISULFID 687 696 By similarity.  
 FT DISULFID 702 755 By similarity.  
 FT DISULFID 807 813 By similarity.  
 FT DISULFID 877 885 By similarity.  
 FT DISULFID 1029 1062 By similarity.  
 FT DISULFID 1066 1073 By similarity.  
 FT CARBOHYD 100 100 N-linked (GlcNAc. .)  
 FT CARBOHYD 105 105 N-linked (GlcNAc. .)  
 FT CARBOHYD 112 112 N-linked (GlcNAc. .)  
 FT CARBOHYD 217 217 N-linked (GlcNAc. .)  
 FT CARBOHYD 317 317 N-linked (GlcNAc. .)  
 FT CARBOHYD 341 341 N-linked (GlcNAc. .)  
 FT CARBOHYD 402 402 N-linked (GlcNAc. .)  
 FT CARBOHYD 418 418 N-linked (GlcNAc. .)  
 FT CARBOHYD 459 459 N-linked (GlcNAc. .)  
 FT CARBOHYD 531 531 N-linked (GlcNAc. .)  
 FT CARBOHYD 698 698 N-linked (GlcNAc. .)  
 FT CARBOHYD 747 747 N-linked (GlcNAc. .)  
 FT CARBOHYD 779 779 N-linked (GlcNAc. .)  
 FT CARBOHYD 820 820 N-linked (GlcNAc. .)  
 FT CARBOHYD 839 839 N-linked (GlcNAc. .)  
 FT CARBOHYD 882 882 N-linked (GlcNAc. .)  
 FT CARBOHYD 907 907 N-linked (GlcNAc. .)  
 FT CARBOHYD 938 938 N-linked (GlcNAc. .)  
 FT CARBOHYD 965 965 N-linked (GlcNAc. .)  
 FT CARBOHYD 973 973 N-linked (GlcNAc. .)  
 FT CARBOHYD 1007 1007 N-linked (GlcNAc. .)  
 FT CARBOHYD 1084 1084 N-linked (GlcNAc. .)  
 FT CARBOHYD 1103 1103 N-linked (GlcNAc. .)  
 FT CARBOHYD 1114 1114 N-linked (GlcNAc. .)  
 FT STRAND 172 178 N-linked (GlcNAc. .)  
 FT HELIX 186 197  
 FT TURN 198 199  
 FT TURN 204 205  
 FT STRAND 208 214  
 FT STRAND 218 222  
 FT TURN 224 226  
 FT HELIX 230 238  
 FT TURN 239 239  
 FT HELIX 250 259

PT TURN 260 261  
 FT HELIX 264 266  
 PT TURN 267 267  
 FT TURN 270 271  
 FT STRAND 274 280  
 FT TURN 285 286  
 FT HELIX 287 289  
 FT TURN 290 299  
 FT TURN 300 301  
 FT STRAND 302 309  
 FT HELIX 311 315  
 FT TURN 316 317  
 FT HELIX 321 328  
 FT TURN 329 330  
 FT HELIX 335 338  
 FT STRAND 339 342  
 FT HELIX 346 350  
 FT HELIX 352 359  
 SQ SEQUENCE 1180 AA; 130808 MW; 8E5DA2BE02362EE4 CRC64;  
 Query Match 31.0%; Score 1894.5; DB 1; Length 1180;  
 Best Local Similarity 36.5%; Pred. No. 5.6e-126;  
 Matches 436; Conservative 221; Mismatches 461; Indels 77; Gaps 24;  
 QY 13 LVFLTGLCSPLNLDHHPRLFPGPPEAEFGYSVLQHVGGQORWMLVGPAPWDPSPGDRGD 72  
 DB 19 LTVLGFCVSNFVDVKNMSFGSGVEDMFGYTVQOQYENEEGKWLIGSPLVGVQPKARTGD 78  
 QY 73 VYRCPVGGAHNAPCAKAGHLGDYQLGNSHSPAV-----NMHLGMSLLETGDCGCFMACAPL 127  
 DB 79 VYKCPVGRERAMPVCVLDLP-----VNTSIPNVTEIKENMTFG-STLVNPNNGGFLACGPL 133  
 QY 128 WSRACGSSVSFGICARVDASFPQGSGLAPTAQRCPTVMVIVLDCGNSIYPWSEVQTF 187  
 DB 134 YAYRCGHLHTVTGICSDVSPTFQVNSFAP-VQECSTQLDIVLVDGNSIYPWSEVIAF 192  
 QY 188 LRLVGLKFLDPQIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRREGRETAKTAQ 247  
 DB 193 LNDLLKEMDIPKQTOQVGIQVQGENVTHEFNLNKYSTEELVVAANKIGRQGGIQTMTAL 252  
 QY 248 AIMVACTGFGSQSHGGRPEARLLVVVTDGSHDGEELPAALKACEAGRVTRYGLAVLGH 307  
 DB 253 GIDTARKEATEARGARRGVKVMVIVTDGSHDNYRLKQVIOCDCEENIORFSIALIGH 312  
 QY 308 YLRQRDPSSFLREIRTIASDPDERFFNVYDEAALTIDVDALGDRIFGLEGSHAENESS 367  
 DB 313 YNRGNLSTEFVBEIKSIASEPTEKHFNFVSDELALVTIVKALGERIFALEATADQSAAS 372  
 QY 368 FGLEMSQIGFSTHRLKDGILFGMVGAYDWGGSVLWLEGGHRLFPFPRMALEDEFPALQNH 427  
 DB 373 FEMEMSQTFSAHYSQDWMVLMGAVGAYDWNGTVVMQKANQMVIPHNTTTFQTE--PAKWE 430  
 QY 428 --AAYLCYSVSSMLLRGGRRLFLSGAPFRHRGVIAFOLKKGAVRVAQSLQGEIOGSY 485  
 DB 431 PLASLYGYTNSATIPGD-VLYIAGQPRYHNTGVVIYKM-EDGNINILQTLGGEIOGSY 488  
 QY 486 FGSELCPLDTRDGTDLVLLVAAPMFLGPQNKETGRVYVYLVGQOSSLTLTGLTLP---- 541  
 DB 489 FGSVLTITIDKDSYTDLLLVGAPMYNGTEKEEGKVYVYAV-NQTRFEYQMSLEPIRQT 547  
 QY 542 -----EPPQDARFGFANGALPDLLNQDGFADVAVGAPLEDHGGQALYLYHG 586  
 DB 548 CCSLKDNSCTKENKNEPCGARFGTAAAVKDLNVDGNDVVI GAPLEDDHAGAVYIYHG 607  
 QY 587 TQSGVRPHPAQRIAAASMPHALSYFGRSVDRGLDDGLVDVAVGAOAGAILLSRPV 646  
 DB 608 SGKTIREAYAQRIPSGGDGKTLKFFGOSIHGEMDNGDGLTDVTIGLGGHALFWARDVA 667  
 QY 647 HLTPSLEVTPQAISVQDRCRRRGOEAVCLTAALCFQVTSRTPGRWDFHQFVRFASLDE 706  
 DB 668 VVKVTMNFEPKNVNIQKKNCRVEKETVCINATMCFHVKLSKEDSIYEADLQYRVTLDS 727  
 QY 707 WTAGARAAFDGSGQRRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALVTVPALDNTTK 766

```
Db 728 LRQISRFPGTQERKIQ--NITVRESECIHRSFYMLDKHDFQDSVRVTLDF---NLTD 782
Qy 767 P--GPVLNEGSPSTIOKLFPFSDGPDNDCVTLVLQVNMDIRGSRKAPFVVRGGRKV 824
Db 783 PEGPVLDDALPNSVHEHIFPAKDCGKERCISDLTLNVST-----TEKSLIIVKSQHDKF 838
Qy 825 LVSTLTENRKENAYNTSLSIIFRNHLHLTP-QRESPIKVECAAPSAHARLCSVGHVP 883
Db 839 NVSLTVNKGDSAYNTRTVQHSPLNLFSGIEBIQKDS-----CESNQNITCRVGYPF 891
Qy 884 FQTGAKVTFLEPFSCSSLLSQVFGKLTASSLSERNGLQENTAGTSAYIQVEPHLP 943
Db 892 LRAGETVTKIIFQFNTSHLSNAIHLSATSDSEEPLESLNDNEVNSISIPVKYEVGLQF 951
Qy 944 SSSTLHRYEHPVGTLP-----VGPGRPKTLRVQNLGCVVSGSLIISALLP-AVAH 996
Db 952 YSSASEHHSVAANETIPEFINSTEDIGNEINVFYIRKRGHPMPDELQUSIFPNUTAD 1011
Qy 997 GGNVFLSLQVITNNAACIVQNLTEPPG-----PPVHPEELQHTNRLNGSNTQCQVVR 1050
Db 1012 GYPVLYPIGSSSDNVNCRPSLEDPFGINSKKMTISKSEVLKRGTIQDCSSTCGVATI 1071
Qy 1051 HLGQLANG-TEVSVGLRLVHNEFFRPAKPSLTIVSTFELGTGEGSVLQLTASRWSES 1109
Db 1072 TCSLLPSDLQVNVSL--LLWKPTFIRAFPSLMLTLRGELKGENSS-LTSSSNKRREL 1128
Qy 1110 LLEVVPQ-RPLISLWLTIGSLVGLLALLVFLWKLGFPAHKKTPEBEKEE 1163
Db 1129 AIQSKDGLGRVPLWILLISAFAGLLMLLILALWKIGFF---KRPLKKOMEK 1180

RESULT 6
ITAL_HUMAN
ID _ITAL_HUMAN STANDARD; PRT; 1151 AA.
AC P56199;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a).
GN Name=ITGAL;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=93155124; PubMed=8428973;
RA Briesewitz R., Epstein M.R., Marcantonio E.E.;
RT "Expression of native and truncated forms of the human integrin alpha
RL J. Biol. Chem. 268:2989-2996(1993).
CC -I- FUNCTION: Integrin alpha-1/beta-1 is a receptor for laminin and
CC collagen. It recognizes the proline-hydroxylated sequence G-P-P-G-
CC E-R in collagen.
CC -I- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-1
CC associates with beta-1.
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC -I- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -I- SIMILARITY: Belongs to the integrin alpha chain family.
CC -I- SIMILARITY: Contains 7 FG-GAP repeats.
CC -I- SIMILARITY: Contains 1 VWFA domain.
CC -I- DATABASE: NAME=PROV; NOTE=CD guide CD49a entry;
CC WWW=http://www.ncbi.nlm.nih.gov/prov/cd/cd49a.htm".
DR PIR; A45226;
DR PDB; 1QCS; X-ray; A/B=140-331.
DR PDB; 1QCY; X-ray; A=141-333.
DR Genew; HGNC:6134; ITGAL.
DR MIM; 192968;
DR GO; GO:0008305; C:integrin complex; TAS.
DR DR GO; GO:0005518; F:collagen binding; TAS.
DR GO; GO:0007160; P:cell-matrix adhesion; NAS.
```

```
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VWFA; 1.
DR PRINTS; PROL185; INTEGRINA.
DR PRINTS; PRO0453; VWFA; 1.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWFA; 1.
DR 3D-structure; Calcium; Cell adhesion; Glycoprotein; Integrin;
KW Magnesium; Receptor; Repeat; Transmembrane.
FT DOMAIN 1 1113 Extracellular (Potential).
FT TRANSMEM 1114 1136 Potential.
FT DOMAIN 1137 1151 Cytoplasmic (Potential).
FT REPEAT 16 75 ?
FT REPEAT 76 ?
FT DOMAIN 147 360 VWFA.
FT REPEAT 349 404 FG-GAP 3.
FT REPEAT 405 457 FG-GAP 4.
FT REPEAT 459 520 FG-GAP 5.
FT REPEAT 540 599 FG-GAP 6.
FT REPEAT 602 654 FG-GAP 7.
FT CA_BIND 470 478 Potential.
FT CA_BIND 552 560 Potential.
FT CA_BIND 614 622 Potential.
FT SITE 1139 1142 GPPKR motif.
FT DISULFID 54 64 By similarity.
FT DISULFID 660 669 By similarity.
FT DISULFID 750 728 By similarity.
FT DISULFID 780 786 By similarity.
FT DISULFID 850 858 By similarity.
FT DISULFID 1002 1034 By similarity.
FT DISULFID 1037 1044 By similarity.
FT CARBOHYD 46 46 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 72 72 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 77 77 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 84 84 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 189 189 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 289 289 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 313 313 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 374 374 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 390 390 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 432 432 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 504 504 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 671 671 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 720 720 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 752 752 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 812 812 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 855 855 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 880 880 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 887 887 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 911 911 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 938 938 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 946 946 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 980 980 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1045 1045 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1055 1055 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1074 1074 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1085 1085 N-linked (GlcNAc...) (Potential).
FT STRAND 143 150
FT TURN 153 154
FT TURN 158 158
FT TURN 159 159
FT TURN 170 171
FT TURN 174 174
FT TURN 176 177
FT TURN 179 186
FT STRAND 190 194
FT TURN 196 197
FT TURN 202 211
FT HELIX 222 231
```



FT	TURN	232	234	
FT	HELIX	236	238	
FT	TURN	239	239	
FT	TURN	242	243	
FT	STRAND	245	252	
FT	TURN	257	258	
FT	HELIX	259	261	
FT	HELIX	262	271	
FT	TURN	272	273	
FT	STRAND	274	281	
FT	HELIX	283	287	
FT	TURN	288	289	
FT	HELIX	293	302	
FT	HELIX	307	310	
FT	STRAND	311	314	
FT	HELIX	317	323	
FT	HELIX	324	330	
SQ	SEQUENCE	1151 AA; 127837 MW; 6B3F3CIAABF52808 CRC64;		
Query Match				
Best Local Similarity 36.34; Pred. No. 1.9e-125; Length 1151;				
Matches 429; Conservative 226; Mismatches 454; Indels 74; Gaps 22;				
QY	23	FNLDEHHPLRFP	PGPEAEFGSVLQHVGGGQRMWLVGAPWDGPGDRRGDVYRCPVGGAH	82
DB	1	FNVVKNMTFSG	VEDMFGYVQYENEBGKWLIGSPVQGPKNRTGDKYKCPVRGE	60
QY	83	NAPCAKHLGDY	QLGNSHPAV-----NMHLGMSLLETDGDDGPFMACAPLWSRACGSSVF	137
DB	61	SLPCVKLDLP	-----VNTSIPNVTENKMTFG-STLVTPNPGGFLACGPLYAVRCGHLY	115
QY	138	SSGICARVDAS	FQPGQSLATQAQRCPTMYDMVVILDCGNSIYPWSEVQTLRRLVGLFI	197
DB	116	TTGICSDVSTF	QVNVSIAP-VQECSTQLDIVILDCGNSIYPWSDVTAFLNDLLKMDI	174
QY	198	DPEIQVGLVQY	GSPVHWSLGDFTKEEVVRAAKNLSRREGRETKTAQAIMVACTEGF	257
DB	175	GPQTQVGI	VOYGENVTHEFNLNKYSTEEVLVAKKIVQGGQRTWTALGTDTRKEAP	234
QY	258	SQSHGGRPEAR	LAVVVYTDGESHDEELPAALCAACEAGRVTRYGIAVLGHYLRQRDPSS	317
DB	235	TEARGARRGV	KVMVITVDGESHDNHRLKKVIQDCEDENIQRFSAIILGNSYRGNLSTEK	294
QY	318	FLBRTIADPDR	RFNFVNTDEAALTDIVDALGDRIFGLESHAENESSFGLBMSQIGF	377
DB	295	FVEEIKSIASE	TEKHFNFVSDDELALVTIVKTLGERIFALEATADQSAASFEMESQTGF	354
QY	378	STHRLKDGILF	GMVGYDWCGLWLGSHRLFPPRMALDEPPPALQNHAAVLYGYSVSS	437
DB	355	SAHYSQWVLM	GAVGADWNGTVVMQKASQIIIPRNTTFNVESTKKNELPLASYLYTVNS	414
QY	438	MLLRGGRRLF	LSGAPRFRHRGKVIATFQKKDQAVRVAQSLQGEQIGSYFGSELCPDTR	497
DB	415	ATASSGDVYI	AGQPRYNTHTGVIIYRM-EDGNIKILQTLGSGEIGSYFGSILLTTDIDK	473
QY	498	DGTTDLVLLA	PMPQLGQNKETGRVYVYLVCQQSLLTLQGLQP-----	541
DB	474	DSNTDILLVA	PMTGMTEKEEQGVVYVAL-NQTRFQYMSLEPIKOTCCSSRQHNSCTT	532
QY	542	---EPPDARF	GFAMGALPOLNODGFADVAVGAPLEDHOGALYHGTQSGVPRHPAQR	598
DB	533	ENKNEPCAR	FGTAIAVKDLNLFNDIVIGAPLEDHGGVAVIYHSGSKTIKREVAQR	592
QY	599	IAAASMPHAL	SYFGSRVGDRLDGLDVLVDVAVGAQGAAILSSRPVHLTPPSLEVTPQA	658
DB	593	IPSGDGKTLK	FFQOSIHGENDLNGDLTDVTIGLGGALFWERDVAVKVTVNFFPNK	652
QY	659	ISVVQRCRRR	GQAVCLTALCALCFQVTSRTPRWDHOFYMRFTASLEWTAGAADAQGS	718
DB	653	VNIQKKNCH	MEGKETVCINATVCPCEVKLKSKEEDTYEADLQYRVTLDSLRQISRFSFGT	712
QY	719	GQRLSPRLRL	SVGNVTCQLHFFHVLDTSDVLRVALTVFALONTYKP--GPVLNBSGP	776

DB	713	QERKQVR--	NITVRKSECTKHSFYMLDKHDFQDSVRITLDF---NLTPDPENGVLDDSLP	767
QY	777	TSIQKLVPP	SKDCGPDNECVTDVLQVNMDIRSKRKAFFVVRGGRRKVLSTTLENKEN	836
DB	768	NSVHEYIP	FAKDCGKKEKISDLSLHV---ATTEKDLLIVRSQNDKFNVSITVKNTKDS	823
QY	837	AYNTLSII	IFSRNLHLASLTPQRESPIKVECAAPSAHARLCSVGHVPVQTCAGKVTFLLEF	896
DB	824	AYNTRIVH	SPNLVFSGI-----EAIQKDSG--ESNHNITCKVGYFPLRRGEMVTFKLIF	877
QY	897	EFSCSLLS	QVFGKLTASSDSLERNGTLQBNTAQTSAYIQVEPHLLFSSESTLHRYEHP	956
DB	878	QFNTSYL	MENVTVIYLSATSDSEBPPETLSDNVNVISIPVKYEVGLQFVSSASEVHISAA	937
QY	957	YGTLP-----	VGPGEFKTLRVQNLGCYVVSGLIISALLPAVAHGN---YFLSLSQV	1007
DB	938	NETVPEVIN	STEDIGNEINIFYLIRKSGSPMPPELKLISFPNMTSNGYPLYPTGLSS-	996
QY	1008	ITNASCIV	QNLTEP-----PGPPVHPEELOHTNRLNGSNTQCQVVRCHLQAKGTEV	1061
DB	997	-SENANCR	PHIPEDPPFSINGSKQMTTSTDLHKRGTILDCTCKEATITCNLTS-SDISQV	1054
QY	1062	SVGLLRVL	VHNEFFRRRAKSLTVVSTFELGTGEGSVLQLTEASRWSLLEVVQT-RPIL	1120
DB	1055	NVSL--IL	WKPTFIKSYFSSNLNLTIRGELRSENAS-LVLSNNQKRELAIQISKDGLPGR	1111
QY	1121	ISWLIGSV	LGVLGLLALLVFLWLKLGPFPAHKIPPEEKREE	1163
DB	1112	VPLWVILLS	FAGLILLMLLILALWKIGFF---KRPLKKKMEK	1151
RESULT 7				
ID	O42094	PRELIMINARY;	PRT; 1171 AA.	
AC	O42094;			
DT	01-JAN-1998	(TrEMBLrel. 05, Created)		
DT	01-JAN-1998	(TrEMBLrel. 05, Last sequence update)		
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)		
DE	Alphal integrin.			
OS	Gallus gallus	(Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Gizzard;			
RX	MEDLINE=97476270; PubMed=9334246; DOI=10.1074/jbc.272.42.26643;			
RA	Obata H., Hayaashi K., Nishida W., Momiyama T., Uchida A., Ochi T.,			
RA	Sobue K.;			
RT	"Smooth muscle cell phenotype-dependent transcriptional regulation of			
RT	the alphal integrin gene.";			
RL	J. Biol. Chem. 272:26643-26651 (1997).			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).			
CC	-1- SIMILARITY: Belongs to the integrin alpha chain family.			
DR	EMBL; AB000470; BAA23160.1; -.			
DR	PIR; A55348; A55348.			
DR	HSP; P18614; IMHP.			
DR	GO; GO:0008305; C:integrin complex; IEA.			
DR	GO; GO:0007160; P:cell-matrix adhesion; IEA.			
DR	GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.			
DR	InterPro; IPR000413; Integrin_alpha.			
DR	InterPro; IPR002035; VWF_A.			
DR	Pfam; PF00357; FG-GAP; 1.			
DR	Pfam; PF00092; VWA; 1.			
DR	PRINTS; PR01185; INTEGRINA.			
DR	PRINTS; PR00453; VWFADOMAIN.			
DR	SMART; SM00191; Int alpha; 4.			
DR	SMART; SM00327; VWA; 1.			
DR	PROSITE; PS00242; INTEGRIN_ALPHA; 1.			
DR	PROSITE; PS00234; VWF_A; 1.			
KW	Cell adhesion; Integrin; Transmembrane.			

QY	SEQUENCE	1171 AA;	130228 MW;	B505A4D65F09736E	CRC64;
Query Match	30.4%;	Score 1853.5;	DB 2;	Length 1171;	
Best Local Similarity	34.7%;	Pred. No. 4.7e-123;			
Matches	416;	Conservative 236;	Mismatches 464;	Indels 83;	Gaps 21;
QY	13	LVFUTGLCSFNLDEHHPRLPFGPPEAEFGYSVLQHVGGQRMWLVGAPWDGPGSDRRGD	72		
DB	8	LTALPHLIGASVNDVKNMTFSGPLEDMFGVTVOQYENEEGKWVLIGSPLVGQPEKRTGD	67		
QY	73	YYRCPUVGAGNAPCAKCHLGD-YQLGNSHIPAVNMHLGMSLLETDDGGGFMACAPLWSRA	131		
DB	68	YYKCPVGRDQSQCILKUNLPDTSVPMVMEKENMTLGTTLV-TNPXGGFGLACQPLYYAK	126		
QY	132	CGSSVFSSGICARVDASFPQGS LAPTAQRCTYMDVVIVLDGNSIYPMSEVQTFURRL	191		
DB	127	CGRLHYTTGVCNSVSTFETKAVAPSVQECKTQLDIVIVLDGNSIYPMSEVTAFLNSL	186		
QY	192	VGKLFDIPQIQVGLVQYGSPPVHMSGIDGFRTKBEYVRAAKNLSRREGRETTKTAQIMV	251		
DB	187	LRNMIDIPQQTQGVIGVQGVTVHFEYFLNTYSTTEEVDMAALRIQRGGGTQMTALGIDT	246		
QY	252	ACTGEPGSGHGRPEARLLVVVTDGSHDGBELPAALKACAGRVTRYGI AVLGHYLR	311		
DB	247	AREEAFTEAHGARGGVQKVMVIVTDGSHONYRLQEVIDKCEDENIQORFAIALGYSYRG	306		
QY	312	QRDPSSFLRELTIASDPDRFRFPNVTDAAALTDIVDALGDRIFGLRGSHAENESSFGLB	371		
DB	307	NLSTKEKVEELKSIASPTKEHFPNVSDLEALVTIVEALGERIFALDEATTDDQQAASFE	366		
QY	372	MSQIGFTHRLKOGILFGMVGAYDWGSGVLMLEGGHRLFPPRMALDEFPFPAQLQNHAAYL	431		
DB	367	MSQAGFSAHYSDQWMLGAVGAYDWNGTVVMVKDSDISFNDTFRDRHSEKIEPLAAYL	426		
QY	432	GYSVSSMLLGGRRLLFSLGAPRFRHRKVIAPQLKDGAVRVAQSLQGBQIGSYFGSEL	491		
DB	427	GTYVNSALTPEG-VLYIAGOPRYNHTQCVIYKMG-EGREVKVLQRLKEGEQIGSYFGGVIT	484		
QY	492	PLDTRDRGTTDVLVAAPMFLGPQNKETGRVYVLVG-----QQSLTLTQGT-----	538		
DB	485	TIDINRUSFTDLLLVGAPPTMGTEKEGQKVYVIALNKTFEYQMSLEPIKQTCSSPLKH	544		
QY	539	-----LOPEPQDARFGPAMGALPDNLQDGFADVAVGAPLEDGHGALYLYHGQTSGVYRPH	594		
DB	545	DTCKVLKNEPCGARFGTAIAAVKDLNLGDYNDIVIGSPLEDDHGA VYIVHGGNTISKK	604		
QY	595	PAQRIAAASHPALSYGRSDVGRHLDDGDDLDVAVAGAAAILLSRSPRVHLTPSLEV	654		
DB	605	YTRIASGGDGEKVKFFGQS VHGEMDLNDGGLDVTITGGLGGAALPWSRDVAEYVNVSMQF	664		
QY	655	TPQAI SVVDRCDRRRGOEAVCLTAALCFQVTSRTPGRWD-HQFYMRFTASIDEWTAGARA	713		
DB	665	MPKSIINLQQNCQINKKARTICINATICFK--TRUKSKEDIFESSLQYWITLDAQRI SRS	722		
QY	714	AFDGSQRLSPRLRLSGVNVTCBQLHFHVLDTSDYLRPVALTVPFALDNTTKP--GPNVL	771		
DB	723	LFTETHERKMQK--NIITKSECIKHNFYMLDKDPDQDSVKVLEF---NFSDESGPVL	777		
QY	772	NEGSPGISQKLVPYSKDCGPDNECVTDLVLQNMNDIRGSRKAPVVRGGRRKVLVSTTLE	831		
DB	778	DSNLPNHSIBYIPPTKDCGAKNKCISDLALNVKASIAAGDSSSPPIVKSRRNDRFTIQLS	837		
QY	832	NRKENAVNTSLSIIFSRNLHLASLTP-ORESPIKVECAAPSAHARLCSVGHVPVQTCAKV	890		
DB	838	NKQDSAYNTRALVQSPNITPAGIEDIQOS-----CRSNHMITCKVGVPFLUKPBEI	890		
QY	891	TFLLEFBFSCSLLSQVFGKLTASSDSLSLRNGTITQENTACTSAIYQEPHLLFSSESTLH	950		
DB	891	SFKISFPQNASYLLENATVDVYATSDSEEPETLRDRNGQVTIPVKVEVGLIYFVSVPKEH	950		
QY	951	RYEVHPYGTLPVG-----PGPEPKTTLRVQNLGCVYVSGVLIISALLPVAHVGN---YF	1001		
DB	951	HVIAAANETIPTAINTTEQIGDEVTLHYRIEKGSHFFMPNLTLLQILYPDVDTTAKNTLLYI	1010		

Qy	1002	LSISQVITNNASC-----IVQNLTPEPGPPVHPBELQHTNRLNGSNTQ	1044
Dd	1011	TTLSH--SQNAICKSSYPVDHLKIGSGKSVPKPKEPTKDTI-----MECDTFS	1058
Qy	1045	QCVVRCHLQLAGKTEVSVCLLRLVHNEFRRAKFSLTVSTFELGTBEGSVLQLTEAS	1104
Dd	1059	CASINCALAP-SDISQNVSL--RVWKPTTIKASHLSLTLVVKALLURSENSLIILRNDHQ	1115
Qy	1105	RWSESLEVVQTRPILITSIMLGSLVGLGILLIALLVFCILKWLGFFFAHKKIPDEEKREE	1163
Dd	1116	KLETMIKISKEPGGSPGVPLWVIPSIFAGLIIILAIIPALKWAGFF---KRPLKKQWEK	1171
 RESULT 8 ITA2 HUMAN			
ID	_ITA2_HUMAN	STANDARD; PRT; 1181 AA.	
AC	P17301;		
DT	01-AUG-1990	(Rel. 15, Created)	
DT	01-AUG-1990	(Rel. 15, Last sequence update)	
DT	25-OCT-2004	(Rel. 45, Last annotation update)	
DE	Integrin alpha-2 precursor [Platelet membrane glycoprotein Ia] (GP1a)		
DE	[Collagen receptor] (VLA-2 alpha chain) (CD49b).		
GN	Name=ITGA2;		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
NCBI_TaxID	=9606;		
[1]	SEQUENCE FROM N.A., AND SEQUENCE OF 30-44.		
RP	TISSUE=Endothelial cells;		
RC	MEDLINE=89308879; PubMed=2545729; DOI=10.1083/jcb.109.1.397;		
EX	Takada Y., Hemler M.E.;		
RA	"The primary structure of the VLA-2/collagen receptor alpha 2 subunit		
RT	(platelet GP1a): homology to other integrins and the presence of a		
RT	possible collagen-binding domain.";		
RL	J. Cell Biol. 109:397-407(1989).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Rieder M.J., Armel T.Z., Carrington D.P., Ozuna M., Kuidanek S.A.,		
RA	Rajkumar N., Toth E.J., Yi Q., Nickerson D.A.;		
RT	"SeattlesNPAs. NHLBI HL66682 program for genomic applications, UW-		
RT	FHCRC, Seattle, WA (URL: http://pga.gs.washington.edu)";		
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 168-368.		
RX	MEDLINE=98019223; PubMed=9353312; DOI=10.1074/jbc.272.45.28512;		
RA	Emaley J., King S.L., Bergelson J.M., Liddington R.C.;		
RT	"Crystal structure of the I domain from integrin alpha2beta1.";		
RL	J. Biol. Chem. 272:28512-28517(1997).		
RN	[4]		
RP	VARIANT GLU-534.		
RP	MEDLINE=94043762; PubMed=7901236;		
RX	Santoso S., Kalb R., Walka M., Kiefel V., Mueller-Eckhardt C.,		
RA	Newman P.J.;		
RT	"The human platelet alloantigens Br(a) and Br(b are associated with a		
RT	single amino acid polymorphism on glycoprotein Ia (integrin subunit		
RT	alpha 2).";		
RL	J. Clin. Invest. 92:2427-2432(1993).		
RN	[5]		
RP	VARIANT GLU-534.		
RX	MEDLINE=20206009; PubMed=10744142;		
RA	Kroll H., Gardemann A., Fechter A., Heberbosch W., Santoso S.;		
RT	The impact of the glycoprotein Ia collagen receptor subunit Al648G		
RT	gene polymorphism on coronary artery disease and acute myocardial		
RT	infarction.";		
RL	Thromb. Haemost. 83:392-396(2000).		
CC	-I- FUNCTION: Integrin alpha-2(beta-1 is a receptor for laminin,		
CC	collagen, collagen C-propeptides, fibronectin and E-cadherin. It		
CC	recognizes the proline-hydroxylated sequence G-P-G-E-R in		
CC	collagen. It is responsible for adhesion of platelets and other		
CC	cells to collagens, modulation of collagen and collagenase gene		
CC	expression, force generation and organization of newly synthesized		

extracellular matrix.

-1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-2 associates with beta-1. Interacts with HPSS.

-1- SUBCELLULAR LOCATION: Type I membrane protein.

-1- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins with I-domains do not undergo protease cleavage.

-1- POLYMORPHISM: Position 534 is associated with platelet-specific alloantigen HPA-5 (Br). HPA-5A/Br(a) has Lys-534 and HPA-5B/Br(b) has Glu-534. HPA-5B is involved in neonatal alloimmune thrombocytopenia (NAIT or NATP). The Lys-534-Glu polymorphism may play a role in coronary artery disease (CAD).

-1- SIMILARITY: Belongs to the integrin alpha chain family.

-1- SIMILARITY: Contains 7 FG-GAP repeats.

-1- SIMILARITY: Contains 1 VWFA domain.

-1- DATABASE: NAME=PROM; NOTE=CD guide CD49b entry; WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd49b.htm".

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

-----

EMBL; X17033; CAA34894.1; --

DR EMBL; AF512556; AAM34795.1; --

DR PIR; A33998; A33998.

DR PDB; 1A0X; X-ray; A/B=166-368.

DR PDB; 1DZ1; X-ray; A=171-355.

DR Genew; HGNC:6137; ITGA2.

DR MIM; 192974; --

DR GO; GO:0008305; C:integrin complex; TAS.

DR GO; GO:0005886; C:plasma membrane; TAS.

DR GO; GO:0005518; F:collagen binding; TAS.

DR GO; GO:0007596; P:blood coagulation; TAS.

DR GO; GO:0007160; P:cell-matrix adhesion; TAS.

DR GO; GO:0009887; P:organogenesis; TAS.

DR InterPro; IPR00413; Integrin\_alpha.

DR InterPro; IPR02035; VWF\_A.

DR Pfam; PF01839; FG-GAP; 2.

DR Pfam; PF00357; Integrin\_alpha; 1.

DR Pfam; PF00092; VWA; 1.

DR PRINTS; PRO1185; INTEGRIN.

DR PRINTS; PRO0453; VWFADOMAIN.

DR SMART; SM00191; Int alpha; 5.

DR SMART; SM00327; VWA; 1.

DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.

DR PROSITE; PS00234; VWF\_A; 1.

KW 3D-structure; Calcium; Cell adhesion; Direct protein sequencing; Glycoprotein; Integrin; Magnesium; Platelet; Polymorphism; Receptor; Repeat; Signal; Transmembrane.

KW Repeat; Signal; Transmembrane.

FT SIGNAL 1 29 Integrin alpha-2.

FT CHAIN 30 1181 Extracellular (Potential).

FT DOMAIN 30 1132 Potential.

FT TRANSMEM 1133 1154 Cytoplasmic (Potential).

FT DOMAIN 1155 1181 Interaction with HPSS.

FT DOMAIN 1155 1161 Interaction with HPSS.

FT REPEAT 45 103 FG-GAP 1.

FT REPEAT 104 78 FG-GAP 2.

FT REPEAT 188 378 VWFA.

FT REPEAT 379 433 FG-GAP 3.

FT REPEAT 434 486 FG-GAP 4.

FT REPEAT 488 549 FG-GAP 5.

FT REPEAT 551 610 FG-GAP 6.

FT REPEAT 615 667 FG-GAP 7.

FT REPEAT 499 507 Potential.

FT CA BIND 563 571 Potential.

FT CA BIND 627 635 Potential.

FT SITE 1157 1161 GFFKR motif.

FT DISULFID 83 92 By similarity.

FT DISULFID 680 737 By similarity.

FT DISULFID 789 795 By similarity.

FT DISULFID 865 876 By similarity.

FT DISULFID 1019 1050 By similarity.

FT DISULFID 1055 1060 By similarity.

FT CARBOHYD 105 105 N-linked (GlcNAc. . .) (Potential).

FT CARBOHYD 112 112 N-linked (GlcNAc. . .) (Potential).

FT CARBOHYD 343 343 N-linked (GlcNAc. . .) (Potential).

FT CARBOHYD 432 432 N-linked (GlcNAc. . .) (Potential).

FT CARBOHYD 460 460 N-linked (GlcNAc. . .) (Potential).

FT CARBOHYD 475 475 N-linked (GlcNAc. . .) (Potential).

FT CARBOHYD 699 699 N-linked (GlcNAc. . .) (Potential).

FT CARBOHYD 1057 1057 N-linked (GlcNAc. . .) (Potential).

FT CARBOHYD 1074 1074 N-linked (GlcNAc. . .) (Potential).

FT CARBOHYD 1081 1081 N-linked (GlcNAc. . .) (Potential).

FT VARIANT 534 534 K->E (in alloantigen HPA-5B; dBSNP:1801106).

FT FTID=VAR\_003977.

FT TURN 170 171

FT STRAND 173 180

FT TURN 183 184

FT HELIX 188 199

FT TURN 200 201

FT STRAND 204 204

FT TURN 206 207

FT STRAND 209 216

FT STRAND 220 224

FT TURN 226 228

FT HELIX 232 240

FT TURN 241 241

FT HELIX 252 262

FT TURN 263 264

FT HELIX 266 268

FT TURN 269 269

FT STRAND 275 282

FT HELIX 289 291

FT HELIX 292 301

FT TURN 302 303

FT STRAND 304 311

FT HELIX 313 317

FT TURN 318 319

FT HELIX 323 330

FT TURN 331 332

FT HELIX 337 340

FT STRAND 341 344

FT HELIX 347 353

FT HELIX 354 362

FT TURN 363 363

SQ SEQUENCE 1181 AA; 129295 MW; 7E1B7ED968A94070 CRC64;

Query Match 28.8%; Score 1756; DB 1; Length 1181;

Best Local Similarity 34.3%; Pred. No. 4.5e-116;

Matches 418; Conservative 219; Mismatches 455; Indels 126; Gaps 31;

Qy 11 LPLVLTGL-----CSPFNLDHHPRLPPGPEAEFGYSVLQHVGGQRWMLVGAPWD 63

Db 11 LPLLVLLALSQGLNCLAYNGLPEAKI PGGPSEQFGYAVQOFINPKGNWLLVGSWS 70

Qy 64 GPSGDRRGDYRCPVGGAHNAPCAKHL-GDYQLGNSHSPAVNMHGLMSLLETDCDGF 122

Db 71 GFPENRMGDVVKCPV-DLSTATCEKLNLTSTSPNVTEMKTNMSLGLLTRNMTGGFL 129

Qy 123 ACAPLWSRACSSVFFSSGICARVDASFPQGLSLAPTAQRCPYMDVIVLDGNSIYPWS 182

Db 130 TCGPLWAQCCGNQYITTVGVCSDISPDLQSASFSPATPCPCLSDIVVVCDESNIYPWD 189

Qy 183 EVQTFELRLVGLKFLDPQIQVGLVQVGPVHWSLGDPRTEKEVRAAKNLSRREGRE 242

Db 190 AVKNFLKRVQGLDIGNTKVGLIQYANPRVFNLTNTYKREMI VATSTQYGGDL 249

Qy 243 TKTAQALMVACTEGFSQSHGGRPEARLLVVVTDGESHGDELPALAKACEAGRVTYGI 302

Db 250 TMTFGALQYARKAYSAAAGRRSATKVMVVVTDGESHGSMKXAVIDQCNDHNLRFGI 309

Qy 303 AVLGHYLRDRDPSSFLREINTIASDPDRFFNVNTEAALTDIVDALGDRIFGLEGSHA 362

Db 310 AVILGLNRLNLDTKNLIKAIASIPTEFYFNVSDAAALKEKAGTGLQIFSIETG-V 368  
Qy 363 ENESSFGLMSQIGFST--HRLKDGILFGMVGAYDWCSSVLM-LEGHRLFPFRMALEDE 419  
Db 369 QGGDNFQMSQVGSFADYSQNDILMLGAVGAFGWSGTTIVQKTSHGHLIFP-----KQA 423  
Qy 420 FPAALQ--NHAAYLVGSVSMILRGRRRLFLSGAPRPHRGKVIKFKKDGAVRVAQSL 477  
Db 424 FQILQDRNHSSYLVGSVAA-ISTGESHFVAGAPRANYTGQIVLVSVNENGHITVIAH 482  
Qy 478 QGBOIGYFSGELCPLDTRDGTDLVLLVAAPLPGQNKETGRVYVY-----LVGQOQL 532  
Db 483 RGDQIGSYFSGVLCSDVDKDTITDVLVGAAPYMSDLKKEGRVYLFITKKGILGQHOF 542  
Qy 533 LTQGTLOPPPODARFGFAMGALPDNLQDGPADVAGAPLEDHCHQALYLXHTQSGVR 592  
Db 543 --LEG---PGIENTRFGSAIALSDINMGDFNDVIVGSPLENQNSGAVIYNGHQHTI 597  
Qy 593 PHPAQRIAAA--SMPHALSFGRSVQGRDLDDGDLVDVAVGAGCAAILSSRPVHLTP 650  
Db 598 TKYSQKILGSDGAFRSHLQFGRSLDGYDGLNGDSITDVSIGAFQGVVQLWSQSIADVAI 657  
Qy 651 SLEVTQAIISVQDRCRRRQGEAVCLTAALCFQVTSRTPGRWDHOFYMRFTASLEWTAG 710  
Db 658 EASFTPEKITLVNKAQ-----IILKLCFSAKFR-PTKQNNQVAIVNITLDA---- 704  
Qy 711 ARAAFDGSQGLSPRL-----RLSVGNV-----TCEQLHPHVLDTSDYLPVALTVT 758  
Db 705 -----DGFSSRVTSRGLFKENNERCLQKNVMAQACPEHIIYIQBPSVMSLDLRVD 759  
Qy 759 FALDNTTKPG--PVLNKGSPTSIQKLVFSSKDCPDNCEVTDILVQVMDIRGSRKAPV 816  
Db 760 ISLEN-----PGTSPALEAYSETAKVFSIPFKDCGEDGLCISDLVDYR-QIPAAQEQPFI 815  
Qy 817 VRGRRKVLVSTTLNENKAYNTSLSIIFSRNLHLASLTPQRESPI---KVEC-AAPSA 872  
Db 816 VSNQNKRLTFSVLTKNKRGSAYNTGIVVDFSENLFASFS-----LPVDGTETVCQVAASQ 871  
Qy 873 HARLCSVGHVPFQGAQVTLLEPFSCSSILSQVFGKLTASSDSLERNGLQENTQAOTS 932  
Db 872 KSVACDVGYPALKRQEQVTTINDFNLQNLQNASLSFQALSGESQENKA--DNLVNLK 929  
Qy 933 AYIQYEPHLLFSBSTLHRYEVHPYGTLP-----VGPGPEFTTLRVQNLGCVVVSGLI 986  
Db 930 IPLYIDAEIHLTRSTNINFEISSDGNVPSIVHSFEDVGPKFIFSLKV-TTGSVPVSMAT 988  
Qy 987 ISALLPAVAGGNYFLSLSQVITNNASCIQNLTEPP-----GPPVHPEELQHTNRL 1038  
Db 989 VIIHIPQYTKBNPLMYLTGVQTDKAGDISCNADINPLKIGQTSSTSSVSFKSENFRHTKEL 1048  
Qy 1039 NGSNTQCVQVRCHLGOLAKGTEVSGVLLRLVHNEFFPRAKFSITVVSTPGLTGEESVL 1098  
Db 1049 NCRFASCSNTVCLWKLVHMKGEYFVNVTTIRWNGTFASSTFTQVLTAAABEINTYNEIY 1108  
Qy 1099 QLTEASRWSESLLEWQTRPLISLMI-----LIGSVLGLLILLALLVFLW 1145  
Db 1109 -----VIEDNTVTPLMTMKPDKAEVTVGIISIIAGILLALLVAILW 1154  
Qy 1146 KLGFPAHK-----KIPBE 1158  
Db 1155 KLGFPRKRYEKWTKNPDE 1172

## RESULT 9

ITA2 BOVIN STANDARD; PRT; 1170 AA.  
AC P53710;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 05-JUN-2004 (Rel. 44, Last annotation update)  
DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GP1a)  
DE (Collagen receptor) (VLA-2 alpha chain) (CD49b) (Fragment).

GN Name=ITGA2;  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94193647; PubMed=7511592;  
RA Kanata T., Puzon W., Takada Y.,  
RT Identification of putative ligand binding sites within I domain of  
RT integrin alpha 2 beta 1 (VLA-2, CD49b/CD29).  
RL J. Biol. Chem. 269:9659-9663(1994).  
CC -I- FUNCTION: Integrin alpha-2/beta-1 is a receptor for laminin,  
CC collagen, collagen C-propeptides, fibronectin and E-cadherin. It  
CC recognizes the proline-hydroxylated sequence G-P-P-G-E-R in  
CC collagen. It is responsible for adhesion of platelets and other  
CC cells to collagens, modulation of collagen and collagenase gene  
CC expression, force generation and organization of newly synthesized  
CC extracellular matrix.  
CC -I- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-2  
CC associates with beta-1.  
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -I- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins  
CC with I-domains do not undergo protease cleavage.  
CC -I- SIMILARITY: Belongs to the integrin alpha chain family.  
CC -I- SIMILARITY: Contains 7 FG-GAP repeats.  
CC -I- SIMILARITY: Contains 1 VWFA domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L25886; AAB59255.1; -;  
DR PIR; I45914; I45914.  
DR HSSP; P17301; IAOX.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF01839; FG-GAP; 2.  
DR Pfam; PF00357; Integrin\_alpha; 1.  
DR PRINTS; PR00092; VWF; 1.  
DR PRINTS; PR01185; INTEGRINA.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00191; Int\_alpha; 5.  
DR SMART; SM00327; VWF; 1.  
DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
DR PROSITE; PS50234; VWFA; 1.  
KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Platelet;  
KW Polymorphism; Receptor; Repeat; Signal; Transmembrane.  
FT NON TER 1 1 Potential.  
FT SIGNAL <1 18 Integrin alpha-2.  
FT CHAIN 19 1170 Extracellular (Potential).  
FT DOMAIN 19 1121 Potential.  
FT TRANSMEM 1122 1143 Cytoplasmic (Potential).  
FT DOMAIN 1144 1170 FG-GAP 1.  
FT REPEAT 34 92 FG-GAP 2.  
FT REPEAT 93 ? VWFA.  
FT DOMAIN 177 367 FG-GAP 3.  
FT REPEAT 368 422 FG-GAP 4.  
FT REPEAT 423 475 FG-GAP 5.  
FT REPEAT 477 538 FG-GAP 6.  
FT REPEAT 540 599 FG-GAP 7.  
FT REPEAT 604 656 Potential.  
FT CA\_BIND 488 496 Potential.  
FT CA\_BIND 552 560 Potential.  
FT CA\_BIND 616 624 Potential.  
FT SITE 472 474 Cell attachment site (Potential).  
FT SITE 1146 1150 GPFKR motif.  
FT DISULFID 72 81 By similarity.

FT DISULFID 669 726 By similarity.  
 FT DISULFID 778 784 By similarity.  
 FT DISULFID 854 865 By similarity.  
 FT DISULFID 1008 1039 By similarity.  
 FT DISULFID 1044 1049 By similarity.  
 FT CARBOHYD 94 94 N-linked (GlcNAc...) (potential).  
 FT CARBOHYD 101 101 N-linked (GlcNAc...) (potential).  
 FT CARBOHYD 332 332 N-linked (GlcNAc...) (potential).  
 FT CARBOHYD 421 421 N-linked (GlcNAc...) (potential).  
 FT CARBOHYD 449 449 N-linked (GlcNAc...) (potential).  
 FT CARBOHYD 464 464 N-linked (GlcNAc...) (potential).  
 FT CARBOHYD 688 688 N-linked (GlcNAc...) (potential).  
 FT CARBOHYD 748 748 N-linked (GlcNAc...) (potential).  
 FT CARBOHYD 945 945 N-linked (GlcNAc...) (potential).  
 FT CARBOHYD 1063 1063 N-linked (GlcNAc...) (potential).  
 FT CARBOHYD 1070 1070 N-linked (GlcNAc...) (potential).  
 FT VARIANT 580 580 G -> V.  
 FT VARIANT 588 588 R -> K.  
 FT VARIANT 725 725 R -> S.  
 SQ SEQUENCE 1170 AA; 128929 MW; ECEPIC5F2448FB1 CRC64;  
 Query Match 28.7%; Score 1754; DB 1; Length 1170;  
 Best Local Similarity 33.9%; Pred. No. 6.1e-116;  
 Matches 411; Conservative 235; Mismatches 465; Indels 102; Gaps 30;  
 QY 9 LFLPLVLTGL---CSPFNLDHHPRLPPGPEAEFGYSVLQVGGGQRMVLGAPWDGP 65  
 DB 2 LQLVLVSQGLNCVAYNVLGPKAKIFSGSPSEQFGYAVQOFTNPKGNMLLVGSPWSGF 61  
 QY 66 SGDRGRGVYRCVPGGAHNPACAKGHL--GDYQLGNSSHPVNMHGMISLLETGDPGGFMAC 124  
 DB 62 PKRMGVDYKCPV--DLSTTTCYKLNLTSTMSMNTKMSLGLTLTRNVGTGGFLTC 120  
 QY 125 APLWSRACGSSVSSGICARVDASFPQGS LAPTAQRCPTVMVIVLDGNSNIYPMSEV 184  
 DB 121 GPLWAQCGSQYITGVCSDVSPDFQLRTSPAPAVQTCPSFIDVVVCDNSNIYPMNAV 180  
 QY 185 QTFRLRLVGKLFIDPEQIQVLGVGSPVHWSLGDPRFKYKEVVRRAAKNLSRREGRETK 244  
 DB 181 KNFLEKEVQGLIDIGFTTQTMGLIQYANNPRVFNLTFSKDEMIKATSTQTFYGGGLTN 240  
 QY 245 TQAQIMVACTEGFSQSHGGRPEARLLVWVTDGSHDGEELPAALKAACEAGRVTRYGIAY 304  
 DB 241 TFAQIQAIVARTASTAGRGATGMVWVVDGSHDGSKLKAVIDQCNKNIILRFGIAV 300  
 QY 305 LGHYLRQRDPSSFLREIIRTASDPDRFFNFVDEAALTDIVDALGRIFGLRGSHAEN 364  
 DB 301 LGYLNRNALDTKNLIKEIKAIASIPTEHFNVSDEADLLEKACTIGEIQIFISIGT-VQG 359  
 QY 365 ESSFGLMSQIGFSTHRLKOG--ILFGVGAIDYGGVSLM--LEGHRLFPFRMALEDEFP 421  
 DB 360 GDNFQEMSQVGSFAEYSPQNNILMLGAVGAYDSGTVMQKTPHGLIFS-----KQAFE 414  
 QY 422 PALQ--NHAALVGVSVSMLLGRRRLFLSGAPFRHRGVIAFQLKDKGAVRVAOSLOG 479  
 DB 415 QILQDRHSSVLYGVSVAS--ISTGNSVHFVAGAPRANTYTGQVLVSVNENVTVIQSQRG 473  
 QY 480 EQIGSYFGSELCPDTRDGTDTLLVAAPFLGPQNKETGRVYVYLVGQOSLLTLQGTLL 539  
 DB 474 DQIGSYFGSVLCVAVDNKDTITDVLVAGAPMYMNDLKEEGRVYLFITKGLNWHQFLE 533  
 QY 540 QPEPPQDARFGAMGALPDNLQDGFADVAGAPLEDGHQGALYLYHGTQSGVRPHPAQRI 599  
 DB 534 GPNGLNARFGSAATAALSDINMDGFNDVIVGSPLENQNSGAVIYNGHEGMIRLRYSQKI 593  
 QY 600 AAASMPHA--LSYFGSVGELDLDGDDLDVAVAGCAQGAAILLSRPIVHLTPSPLETPQ 657  
 DB 594 LGSDFRFSHLLQYFGSLDGYDGLNGDSITDVSVGAFQGVVQLVQSADIADVSADSTPK 653  
 QY 658 AISVQQRDCRRRQGEAVCLTAALCFQVTSRTPGWDHGFYVNRFTASIDEMTAG----ARA 713  
 DB 654 KITLLNKAE-----IKLKLCSAKFR--PTQNQNAIVNITIDEDQFSSRVISRG 704

QY 714 AFDGSGORLSRRLRSVGNVTCQQLHFHVLDTSDYLRPVALTVPFALDNTTKPG--PVL 771  
 DB 705 LFKENNERCLOKTMIVSQAQ--RCSEYIIHQEPIISPLNLCMNISLEN---PQTNPAL 760  
 QY 772 NEGSPSTIQKLVPFKDCGPDNECVTLQVQNMNDIRGSRKAPFVWGRGRKVLVSTTLE 831  
 DB 761 EAYSETVKVFSIPPHKDCGDDGVCSIDLVLNV--QQLPATQQOQPFIVSNQNKRLTFSVLK 819  
 QY 832 NRKENAYNTSLTIFSRNLHLASLTPORESPI---KVEC--AAPSAHARLCSVGHVPQTG 887  
 DB 820 NKKEAYNTSIVVDVFSENLFFASWS---MPVDGTEVTCQIASQSKSVTCNVGYPALKSK 875  
 QY 888 AKYTELEFFSCSSLLSLOVFGKLTASSDSLSRNGTLQENTAOISAYIQVPHLLFSSES 947  
 DB 876 QQVTFINFDNQLNQNASISPRALSESQEE--MADNSVNLKSLLYDAEIHITRST 933  
 QY 948 TLHRYE-----VHPYGTLPVGPGBFKTLRVQNLGCVYVVGSLIISALLPAVAHG 997  
 DB 934 NINFYEVSLDGNVSSVHVSFEDI---GPKFIFSIKV--TTSVVPVSMASVIIHIPOYTKD 988  
 QY 998 GNVFLSLSQVITNNA---SCIVQNLTEPPGPPVHP-----BELQHTNRLNG 1040  
 DB 989 KNPLMYLTGVHTDQAGDISCEAE-----INPLKIGQTSSTSSVSKSENFRHIKELNC 1039  
 QY 1041 SNTQCQVVRCHLGOLAKGTEVSGLLRLVHNEFPRAKFKSLTVVSTFELGTEESGVLOL 1100  
 DB 1040 RTASCSNIMWLRLDQVKGEYFLNVSTRWNGTFAASTFQTVQLTAAAEIDTYNPQIYVI 1099  
 QY 1101 TEASRWSESLEVVQTRP---ILISLWILGSLVGLLLALLVFLCWLKLGFFPAHK---- 1153  
 DB 1100 EE-----NTVPIPLTMKPEKVEVPTGVIVGSIAGILLALLVAILKWLKGLFFRKYKEM 1155  
 QY 1154 -KTIPEEKREKL 1165  
 DB 1156 AKNPDETDETEL 1168  
 RESULT 10  
 ITA2 MOUSE  
 ID ITA2 MOUSE STANDARD; PRT; 1178 AA.  
 AC Q62469; Q62163;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Integrin alpha-2 precursor (platelet membrane glycoprotein Ia) (GP1a)  
 DE (Collagen receptor) (VLA-2 alpha chain) (CD49b).  
 GN Name=Itga2;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;  
 RX MEDLINE=94363406; PubMed=8081889;  
 RA Edelman J.M., Chan B.M., Uniyal S., Onodera H., Wang D.Z.,  
 RA Damjanovich L., Latzer D.B., Finberg R.W., Bergelson J.M.;  
 RT "The mouse VLA-2 homologue supports collagen and laminin adhesion but  
 RT not virus binding.";  
 RL Cell Adhes. Commun. 2:131-143(1994).  
 RN [2]  
 RP SEQUENCE OF 450-1178 FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=94355691; PubMed=7521231;  
 RA Wu J.B., Santoro S.A.;  
 RT "Complex patterns of expression suggest extensive roles for the alpha  
 RT 2 beta 1 integrin in murine development.";  
 RL Dev. Dyn. 199:292-314(1994).  
 CC -!- FUNCTION: Integrin alpha-2/beta-1 is a collagen receptor, being  
 CC responsible for adhesion of platelets and other cells to  
 CC collagens, modulation of collagen and collagenase gene expression,  
 CC force generation and organization of newly synthesized  
 CC extracellular matrix. It is also a receptor for laminins, collagen

CC C-propeptides and E-cadherin. Mice homozygous for a null mutation  
 CC in the alpha-2 die very early in embryogenesis.  
 CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-2  
 CC associates with beta-1.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins  
 CC with I-domains do not undergo protease cleavage.  
 CC -1- SIMILARITY: Belongs to the integrin alpha chain family.  
 CC -1- SIMILARITY: Contains 7 FG-GAP repeats.  
 CC -1- SIMILARITY: Contains 1 VWFA domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; Z29987; CAA82877.1; -;  
 CC DR EMBL; X75427; CAA53178.1; -;  
 CC DR PIR; S44142; S44142.  
 CC DR HSP; P17301; IAOX.  
 CC DR MGD; MGI:96600; Itga2.  
 CC DR InterPro; IPR000413; Integrin\_alpha.  
 CC DR InterPro; IPR002035; VWF\_A.  
 CC DR Pfam; PF01839; FG-GAP; 2.  
 CC DR Pfam; PF00357; Integrin\_alpha; 1.  
 CC DR Pfam; PF00092; VWA; 1.  
 CC DR PRINTS; PRO1185; INTEGRINA.  
 CC DR PRINTS; PRO0453; VWFADOMAIN.  
 CC DR SMART; SM00191; Int alpha; 5.  
 CC DR SMART; SM00327; VWA; 1.  
 CC DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 CC DR PROSITE; PS0234; VWA; 1.  
 CC DR Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Platelet;  
 KW Receptor; Signal; Transmembrane.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1178  
 FT DOMAIN 27 1129 Integrin alpha-2.  
 FT TRANSFAM 1130 1151 Extracellular (Potential).  
 FT DOMAIN 1152 1178 Potential.  
 FT REPEAT 42 100 Cytoplasmic (Potential).  
 FT REPEAT 101 ? FG-GAP 1.  
 FT REPEAT 185 ? VWA.  
 FT DOMAIN 185 375 VWA.  
 FT REPEAT 376 430 FG-GAP 3.  
 FT REPEAT 431 483 FG-GAP 4.  
 FT REPEAT 485 546 FG-GAP 5.  
 FT REPEAT 548 607 FG-GAP 6.  
 FT REPEAT 612 664 FG-GAP 7.  
 FT CA\_BIND 496 504 Potential.  
 FT CA\_BIND 560 568 Potential.  
 FT CA\_BIND 624 632 Potential.  
 FT SITE 480 482 Cell attachment site (Potential).  
 FT SITE 1154 1158 GFFKR motif.  
 FT DISULFID 80 89 By similarity.  
 FT DISULFID 677 734 By similarity.  
 FT DISULFID 786 792 By similarity.  
 FT DISULFID 862 873 By similarity.  
 FT DISULFID 1016 1047 By similarity.  
 FT DISULFID 1052 1057 By similarity.  
 FT CARBOHYD 102 109 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 109 109 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 429 429 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 457 457 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 472 472 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 696 696 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1054 1054 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1071 1071 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1078 1078 N-linked (GlcNAc... ) (Potential).  
 SQ SEQUENCE 1178 AA; 128926 MW; 1F194B9C0240F465 CRC64;

Query Match

28.6%; Score 1744; DB 1; Length 1178;

Best Local Similarity 35.3%; Pred. No. 3.2e-115;  
 Matches 420; Conservative 221; Mismatches 479; Indels 70; Gaps 31;  
 QY 9 LFLPLVFLT-GL--CSPENLDEHHPLFPQPEAERGYSVLQVGGQRMVLMVGADWGP 65  
 DB 10 LLLQLMLVQGIILCLAYNVGLPGAKIFGPGSEQFGYSVQQLTNPQGNWLLVGSWSGF 69  
 QY 66 SGDRRGDVYRCPVGGAAHNPACAKGHL-GDYQLGNSSHPAVNMHLGMSLLETDCDGGFMAC 124  
 DB 70 PENMGDVYKCV-DLPTATCEKLNQNSASISNVTETKTNMSLGLTLTENPGTGGFLTC 128  
 QY 125 APWLRACSSVSSGICARVDASPOQSLAPTACRCPTMYMDVIVLVDGSNSIYPWSEV 184  
 DB 129 GPLWAHQCGNYATGICSDVSPDFQLTSFSPAVQACPSLVDVVVVCDESNIYPWEAV 188  
 QY 185 QTLRLRLVGLKFLDPQIQLVGLVOYGESPVHWSLGDFTKEBVVPAAKNLSRREGTEK 244  
 DB 189 KNFLVFPVTGLDIGPKKTQVALIQYANSPRIIPLNLDFTKEDMVOATSETROGGDLTN 248  
 QY 245 TAAQIMVACTEGFSQSHGGRPEAARLLVVVTGESHGDEELPAALKACAGRVTRYGIIV 304  
 DB 249 TFAIEFARDYAYQSISGGRPGATKVMVVVTGESHGSKLKTIVIOQDNDEILRGLIV 308  
 QY 305 LGHYLRQRDPSPFLREIRTIASDPDERFFNVYTDAAITDIVDALGDRIFLGECSHAEN 364  
 DB 309 LGYLNRNALDTKNLKEIKAIASPTERYFFNVVADEAALLEKAGTLGEQIFSTEGT-VQG 367  
 QY 365 ESSFGLEMSQIGFSTHRL--KQILFGMVGYDMGWSVLWLEGHR--LFPFRMALEDEF 420  
 DB 368 GDNFQEMAAQVGSADYAFONDILMLGAVGAFDWSGTLV-QETSHKPVFPF-----KQAF 421  
 QY 421 PPAIQ--NHAAYLGYSVSMLLRGGRRLFLSGAPRFRHRGKVIATFOLKXDGAVRVASQLQ 478  
 DB 422 DQVLQDRNHSSFLGYSVAALSTEDGVH-FVAGAPRANYTGOIVLYSVNKGNTVTIQSHR 480  
 QY 479 GEIGSYFSGELCPLDTRDGTDDLVAAPMFLGPONKETGRVYVYLVGQQSLTLQGT 538  
 DB 481 GDQIGSYFGSLCSDVDVDKDTITDVLVGAPTYMNDLKEEGKVLFTTKGLNQHQFL 540  
 QY 539 LOPEPQDARRGFAMGALPDINODGFADVAVGAPLEDHGQGLYLVHGTSGVRPHPAQR 598  
 DB 541 EGPEGTGNARFGSAIAALSDINMDGPDVIVGSPVENENSGAVIYNGHQGTIRTKYSQK 600  
 QY 599 IAAA--SMPHALSYFGRSVDGRDLDDGDLVDVAVGAQGAAILLSRPIVHLTPSLEVP 656  
 DB 601 ILGNGAPRRLHQLFFGRLDGYDGLNGDSITDVISIGALQVQLWSQSADVAIEALFTP 660  
 QY 657 QAISVVQDRCRRRQGBAVCLTAALCFQVTSRTPRWDHQPYMRFTASLDEWTAGARAFD 716  
 DB 661 DKITLLNKDAK-----ITLKLCPFRABFRPAGQ-NNQVAILFNMTLDADGHSSRVTSR 711  
 QY 717 GSGQRLSPRLR--LSVGNV-TCEQLHFVHVDTSYLRPVALTVTFTALDNTTKRG--PVL 771  
 DB 712 GVFNRSERFLQKNVNVNEVQKSEHHISIQKPSDVNVNPLDRLVDISLEN---PGTSPAL 768  
 QY 772 NEGSPTSIOKLVPFSKDCGPDNECVTDVLQVNMIDIGRSKAPFVVRGGRKVLVSTLE 831  
 DB 769 EAYSETVKVFSIPFYKECGSDGICISDLILDV-QQLPAIQTSQFIVSNQNKRLTFPSVLK 827  
 QY 832 NRKENAYNTSLSIIPSRNLHLASLTPORESPI---KVECAAPSAHARL-CSVGHVPQTG 887  
 DB 828 NRGESAYNTVLAEPSENLFASFSS---MPVDGTEVTCVSGSSQKSVTCVGVYPALKSE 883  
 QY 888 AKYTELEPEFSCSSLLSGLVFGKLTASSDSLENGTLQENTAQTSAYIOVEPHLLSSES 947  
 DB 884 QQVTFITNFDNFNLQNLQNAAINQAFSESQETNKA--DNSVSLTIPLLDAELHLTRST 941  
 QY 948 TLHRYEVHPYGTLP-----VGPGBFKTTLRVQNLGCVVSGLIISALLPAVAHGNYF 1001  
 DB 942 NINFYEISSDENAPSVIKSVEDIGPKFISLKV-TAGSAPVSMALVTIHIPOYTKENPL 1000  
 QY 1002 LSLSQVITNNA---SCIT--VQNLTEP---PGPVHPPEELQHTNRLNGSNTOCQVVRCHLG 1053



Db 1001 LYLTIOTDQAGDISCTAEINPLKLPHTAPSVSFKNENPRHTKELDCRTTSCSNITCWLK 1060  
Qy 1054 QLAGTETVSGLLRLVNEFFRAKFKSLTVVSTFELGTBEGSVQLQTEASRWSESLLEV 1113  
Db 1061 DLHMAEYFINTVTRVNRTPFAASTFTQVLTAARAEIDTHNPQ-LFVIEENAVTIPLMIM 1119  
Qy 1114 VQTRPILISLWILIGSVLGGILLALLVFCILWKLGF--AHKKI---PEE 1158  
Db 1120 KPTEKAEPVGTGVIIGSIAGILLLLAMTAGLWKLGFPPKQYKMGQNPDE 1169

RESULT 11  
Q6P1C7  
ID Q6P1C7 PRELIMINARY; PRT; 1178 AA.  
AC Q6P1C7  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Integrin alpha 2.  
GN Name:Itga2;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Eye;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Heide F.,  
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
RA Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Vallation D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalska U., Smalls J.D., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Eye;  
RA Strausberg R.;  
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.  
DR EMBL; BC065139; AAH65139.1; -.  
DR HSSP; P17301; 1A0X.  
DR GO; GO:0005615; C:extracellular space; TAS.  
DR GO; GO:0016021; C:integral to membrane; TAS.  
DR InterPro; IPR004113; Integrin\_alpha.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF01839; FG-GAP; 2.  
DR Pfam; PF00357; Integrin\_alpha; 1.  
DR Pfam; PF00092; VWA; 1.  
DR PRINTS; PR01185; INTEGRINA.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00191; Int\_alpha; 5.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
DR PROSITE; PS0234; VWF\_A; 1.  
KW Cell adhesion; Integrin; Transmembrane.  
SQ SEQUENCE 1178 AA; 128954 MW; 62FAEA820242A9B6 CRC64;

Query Match 28.6%; Score 1743.5; DB 2; Length 1178;  
Best Local Similarity 35.2%; Pred. No. 3.5e-115; Indels 69; Gaps 30;  
Matches 417; Conservative 222; Mismatches 477;  
Qy 13 LVFLTGL--CSPNLDHHRPLRFPGPPEAEFGYSVLQHVGGGORMMLVGPWDPGSDRR 70  
Db 15 LMLVQGLNCLAYNVGLPGAKIFSGPSSEQFGYSVQQLTNPQGNWLLVGPSPWSPFENRM 74  
Qy 71 GDVYRCPVGGAHNAPCAKHL--GDYQLGNSSHPAVNMHLGMSLLETDDGDFMACAPLWS 129  
Db 75 GDVYKCPV-DLPTATCEKLNQNSASISNVTETKNNLSGLTLTRNPGTGFGTLCGLWA 133  
Qy 130 RACGSSVFSSGICARVDASPOGSLAPTAORCPYTMVDVIVLDGNSIYPWSSVQTFLR 189  
Db 134 HQCGNQYATGICSDSPDFLTSPSPAVQACPDLVDVVVVVCDSESIYPWEAKNPLV 193  
Qy 190 RLVGKLFIDPEQIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRRREGRETAKAI 249  
Db 194 KFTVGLDIGPKTQVALIQVANEPRIFINLNDRETEDMVQATSETRQHGGLTNTFRAI 253  
Qy 250 MVACTEGFSQSHGGRPEARLLVVVTDGESHDEELPAALKACBAGRTRYGIAVLGHL 309  
Db 254 EFARDVAYSQTSGRPGCATKVMVWVTDGESHGSKLTVIQQCNDDEILRFIAVLGYLN 313  
Qy 310 RRQDPSSFLEIRTIASDPDERFFENVTDEAALTDIVDALGDRIFGLESHAENESSFG 369  
Db 314 RNALDTKNLIKEIKATASTPTEYFFNVADEAALLKAGTLGEOIFSIETG-VGGDNFQ 372  
Qy 370 LEMSGIGFSFHL--KDGILFGVMGAYDWGSGVLMLGEGHR--LFPFRMALEDFPPALQ 425  
Db 373 MEWAQVGSADYAPONDILMLGAVGAFDWSGLV-QETSHKPVIFP-----KQAFDQVLQ 426  
Qy 426 --NHAAYLVYSVSMILRGRRRLFLSGAPRFRHGRKVIAPQLKKDGAVRVAQSLOGSQIG 483  
Db 427 DRNHSSFLGYSVAAISTEDGVH-FVAGAPRANTGQIVLYSVNKGQNVTVIQRHGQIG 485  
Qy 484 SYGSELCLPDTDRDGTITDVLVAAPFLGPONKETGRVYVLYVQGSLLTLOQTLOPEP 543  
Db 486 SYFGSVLCSDVDKDTITDVLVGAAPYMDLKKKEGKVLFTITKGLNQHOHFLGEPG 545  
Qy 544 PQDARFGMAGALPDNLQDGFADVAVGAPLEDHGQALYLYHGTQSGVRPHAPQRIAAA- 602  
Db 546 TGNARFGSATAALSDINMGDFNDVIGSPVENENSGAVIYNGHQITIRKYSQKILGSN 605  
Qy 603 -SMPHALSYPGRSDGLDGLDLDVAVGAQAAILLSSRPVHLTPSLEVTPQAIISV 661  
Db 606 GAFRRHLQFFGRSLDGYDLNGDSITDVSIGALQVQLWSQSIADVAIEALFDPDKITL 665  
Qy 662 VQRDCRRRGOEAVCLFAALCFQVTSRTPGRWDHQFYMRFTASLDDEWTAGARAAFDGSGOR 721  
Db 666 LNKDAK-----ITLKLCFRAEFEPAGQ-NNQVAILFNMTLDADGHSSRVTSGRVPRE 716  
Qy 722 LSPRRLR--LSVGNV-TCEQLHFLVLDTSVLRPVALTVTFALDNTTKPG--PVLNCGSP 776  
Db 717 NSERFLOKNNVMEVQKCEHHISIQKPSDVVNPDLRVDISLEN---PGTSPALEAYSE 773  
Qy 777 TSIQKLVFFKOCQPDNECVTLVLQVNMIDIRSGRKAPFVVRGGRKVLVSTTLLENKEN 836  
Db 774 TVKVFSIPFYKEGSDGICISDILDV-QQLPAIQTSFIVSNQNKRLTFSVLKNGES 832  
Qy 837 AYNTSLSIIFSRNLHLASLPQRESPI---KVECAAPSAHARL-CSVCHPVFOGAKVTF 892  
Db 833 AYNTVLAEEFSENLFASFSS---MPVDGTEVTCEVSSQKSVTCVGYGPAKSEQQVTF 888  
Qy 893 LLEPERSCSILLSQVFCGLTASDSLERNGTLOENTTAQTSAYIOYEPHLLFSSTLHRY 952  
Db 889 TIINFDFNLQNLQAAINFAQFSESQETNKA--DNSVSLTIPLYDLIELHLTRSTNIFY 946  
Qy 953 EVHPHYGTLP-----VGPGEFKTTLRQNLGCVVVSGLIISALLPAVAGGNFYLSLSQ 1006  
Db 947 EISSDENAPSVIKSVEDIGPKFTFSLKV-TAGSAPVSMALVTIHPQVTKENPLLYLTG 1005  
Qy 1007 VITNNA---SCI---VQNLTEP---PGPPVHPEELQHTNRLNGSNSTQCQVVRCHLGQLAKG 1058

```

Db 1006 IQTQAGDISCTABINPLKLPHTAPSVPKNNPRHTKELDCRTSCSNITCWLKDLHMK 1065
Qy TEVSUGLLRLVHNEFFRRKFKSLTVVSTPELGEESVQLQTEASRWSLSLLVQVTRP 1118
Db 1066 AEFYINVTWNTFAASTFTQVQLTAABIDTHNQ-LFVIEENAVTIPLMIMKTEK 1124
Qy 1119 ILISLWILGSLVGLLLALLLVCLMKLGF--AHKKI---PDE 1158
Db 1125 AEPVTVIGSIIAGILLALLAMTAGLWKLGFRRKQYKKMGQNPDE 1169

RESULT 12
ID Q8WY18 PRELIMINARY; PRT; 823 AA.
AC Q8WY18;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE M5TP018.
OS Homo sapiens (Human).
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=Aorta;
RA Xu Y.Y., Sun L.Z., Wu Q.Y., Liu Y.Q., Liu B., Zhao B., Wang X.Y.,
RA Song L., Ye J., Sheng H., Gao Y., Zhang C.L., Zhang J., Wei Y.U.,
RA Sun Y.H., Jiang B.Q., Zhao X.W., Liu S., Liu L.S., Ding J.F.,
RA Gao R.L., Qiang B.Q., Yuan J.G., Liew C.C., Zhao M.S., Hui R.T.;
RA Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
DR EMBL; AF111999; AAL39001.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0005515; F:protein binding; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.
DR Pfam; PF01839; FG-GAP; 1.
DR PRINTS; PR01185; INTEGRA.
DR SMART; SM00191; Int alpha; 4.
DR KEGG; K04401; Integrin; Transmembrane.
KW Cell adhesion; Integrin; Transmembrane.
SQ SEQUENCE 823 AA; 92672 MW; D54E78079DCD4925 CRC64;

Query Match 24.4%; Score 1487.5; DB 2; Length 823;
Best Local Similarity 39.6%; Pred. No. 4e-97;
Matches 327; Conservative 151; Mismatches 308; Indels 39; Gaps 13;

Qy 372 MSQIGFSTRHLKDGILFGMVGAYDVGSGVLWLEGGHRLFPFMALEDEFFPALONHAYL 431
Db 1 MSQIGFSSHVVEDGVLLGAVAYDVGWGAVALKETSAGKVIPLRESYLFKPELKNHGYL 60
Qy 432 GYSVSMLLRGRLFLSGAPFRHRGKVIAFQLKDCQAVRAVQASLOEQIGSYFGSELG 491
Db 61 GYTVTSVSSRQGRVYVAGAFRNHTGKVLFTWHNRSLSLTIHQMRGQQLGSYFGSEIT 120
Qy 492 PLDTRDGTTLVLAAPMFQPNKETGRVYVYLVGQQLSLTLTQGLTQPPPP-ODARFG 550
Db 121 SVDIDGQVTVLLVAGPMYFN-EGRRGKGVYVEL-RQNRVFNVTGLKDSHSYQNAF 178
Qy 551 FAMGALPDNODGFADVAVGAPLEDHCGALYLVHGTQSGVRPHPAORIAASMPHALSY 610
Db 179 SSIAVRDLNDSVNDVVVGAPELNDHAGAIYIFHGFSGSLTKTPKORITASELATGLQY 238
Qy 611 FGRSDGRDLDDGDLVDVAVGACAAIILSSRPVHLTPSLVTPQAISSVVQDCRRRG 670
Db 239 FGCISHGQLDNEGLDLAVGALGNVILWSRPVQVNASLHFPEPSKINIFHRDCKESG 298
Qy 671 QEAVCLTAAALCFQVTSRTPTGRWDHQFVYMRFTASLDIEWTAGARAAPDGGQRLSPRLRLS 730

```

```

Db 299 RDATCAAALCFCTPIIFLPHFQTTTIGIRYNATMDERRYPTRAHLDEGGDRFTNRAVLLS 358
Qy 731 VGNVTCBQLHFHLDTSYLRPVVALTTFALDNTTKPGVNLGSGPTSIOKLVPFSDCG 790
Db 359 SQGLERINFLVLDIADYVKPVTFSVYSLEDDPH-GPMLDDGWTTLVSVFPFNGCN 417
Qy 791 PDNECVTDLVQVMDI-----RGRK-----APFVVRGGRKRVLVST 828
Db 418 EDEHCVPLVDARSDLPTAMEYQORVLRKPAQDCSAYTLSPDTTTFIESTQRVAVEA 477
Qy 829 TLENKRNAYNTSLIIFSRNLHLASLTQRESPIKVECAAPS--AHARLCSVGHVPFQT 886
Db 478 TLENKRNAYSTVINISQSANLQFASLIQKEDSGSIECVNEERRLQKQVCNVSYPFRA 537
Qy 887 GAKVTLLPEFSCSSLSOVFGKLTASSDSLRNGTLOENTATQTSAYIQEYEPHLLFSSE 946
Db 538 KAKVAFRLDFEFSKSLFLHLEIELAAGSDSNERDSTKEDNVAPLPHLYEADVLFTRS 597
Qy 947 STLHRYEHPYGTLP--VGPGEFKTTLRVQNLGCVVSGLIISALLPAVAHGNVPLSL 1004
Db 598 SSLSHYEVKLNSSLERYDYGIGPPFCIFRIQNLGLFPFHGMWMMKITIPIATRSNRLKL 657
Qy 1005 SQVITN--NASC-IVQNLTEPPGPPVHPPELQHTNRLNGSNTQCVVRCHLQLAGTEV 1061
Db 658 RDLTDEANTSCNIGNSTEYRTPVE-EDLRAPQLNHSNDSVVSINCNI-RLVNPQEI 715
Qy 1062 SVGLLRLVHNEFFRRKFKSLTVVSTFELGTGSGVLQLTEASRWSLSLLEVQVTRPILI 1121
Db 716 NFHLGLNMLRSALKYKSMKIMVNAALQRPSPFIFREEDPSRQIVFEISKQEDMQV 775
Qy 1122 SLMLIGSVLGGLLALLVLCWLKLGFPFAHKKIPPEEKREKLE 1166
Db 776 PIWIVGSTUGLLALLVALLVWLKLGFPFARSAR---RRREPGLD 816

RESULT 13
Q8CE84 PRELIMINARY; PRT; 823 AA.
ID Q8CE84;
AC Q8CE84;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
DE library, clone:4732459H24 product:MSTP018 homolog.
GN Name=Itgall;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RX Carninci P., Hayashizaki Y.;
RA "High-efficiency full-length cDNA cloning.";
RA Meth. Enzymol. 303:19-44(1999).
RN 2;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RA "Functional annotation of a full-length mouse cDNA collection.";
RN 3;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RA "Analysis of the mouse transcriptome based on functional annotation of
RA 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN 4;
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata K., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto K., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multiplexed capillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanegaki T., Hara A., Hoshizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsumoto Y., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito K., Satoh K., Sakai K., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
DR EMBL; AK028821; BAC26137.1; -.
DR MGD; MG1:242114; Itgal1.
DR GO; GO:0006929; P:substrate-bound cell migration; IMP.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 1.
DR PRINTS; PR01185; INTEGRIN.
DR SMART; SM00191; Int alpha; 4.
DR Cell adhesion; Integrin; Transmembrane.
SQ SEQUENCE 823 AA; 92264 MW; A330236324A0E089 CRC64;

Query Match 24.1%; Score 1473; DB 2; Length 823;
Best Local Similarity 39.6%; Pred. No. 4.3e-96;
Matches 328; Conservative 145; Mismatches 311; Indels 44; Gaps 12;

Qy 372 MSQIGFSTRLLKDGILFGVCAGYDWGGSVLWLEGGHRLFPFRMALEDFEPALQNHAYL 431
Db 1 MSQIGFSSHVEDGLLGAVGAYDWNGAVLKETSAGKVPHPRESYKLEFPBELKNHAYL 60
Qy 432 GYSVSSMLLRGGRLLFLSGAPFRHRGKVIQAFQKKGCAVRAVQSLQGEIGSYFGSEL 491
Db 61 GYTVTSVSSRQGRVYAGAFRNHTGKVLFSMHNRLTIHQALRGEIGSYFGSEIT 120
Qy 492 PLDTDRGTTDVLVAAPMFQPNKQETGRVYVYLVGQSLTLTQGLQPEPP-ODARFG 550
Db 121 SVDVNDVRDVTDLVVGAPMYFS-EGRERKGVVYVNL-RQNRVYVNGTLKDSHSYQARFG 178
Qy 551 FAMGALPDNODGADVAVGAPLEDHOGALYLVHGTQSGVRPHPAQRIAAASMPHALSY 610
Db 179 SCTASVQDLNODSVNDVVVVGAPLEDHSHRGAIYIFHGFOTNLKPKPMORITASELAPLQH 238
Qy 611 FGRVSDGRLLDGDGLVDVAVGACAAILLSSRPVHLTPSLVTPQAISVVQDCRRRG 670
Db 239 FGCSIHGQLDNEGLDVLAVGALGNVAVLWARPVQINASLHPSPKINFHKDCKRG 298
Qy 671 QEAVCLTAALCFQVTSRTSGRWDHQFYMRFTASLDWTAGARAAAFDGGQRLSPRLRLS 730

```

```

Db 299 RDATCLAAFLCFTPIPIFLAPHQATATGIRNATMDERRYPRAHLDDGGDFTTRAVLLS 358
Qy 731 VGNVTCQLHFHVLDTSDYLRPVALTYTTFALDNTTKPGVLNEGSPITSIQKLVPFSDCG 790
Db 359 SGQEHQCQRINFHVLDTADYVKPVAFSVEYSLEDDPDN-GPMLDNGWPTTLRVSVFFWNGCN 417
Qy 791 PDNECVTDIVLQVNMDIRGRK-----APFVVRGGRKVLVST 828
Db 418 EDEHCVDPDLVDARSDDLPTAMEYQCVLRPAQDCSSYTLTSDFTVTFTIESTRRRVAVEA 477
Qy 829 TLENRKENAYNTSLIIFSRNLHLASLTPORESPIKVECAAPS--AHARLCSGVGHPVPQT 886
Db 478 TLENRGENAVNLINISQSENLPASLIQDDSDNSIECNEERLHKVCNVPFPFRA 537
Qy 887 GAKVTFLLPEFSCSSLLSQVFGKLTASSLSLERNGLTQENTAQTSAYIOYEPHLLFSSE 946
Db 538 KAKVAPRLDFEFSKSVFLHLQIHLGAGSDSHEQDSTADDNTALLRFLHLYEADVLFRS 597
Qy 947 STLHRYEVHPYGTLPV--GPGPEKFTLRVQNLGCYVVGSLIIISALLPAVAHGNYFLSL 1004
Db 598 SSLSHFEVKANSLSLESDGIGPPFCVFKVQNLGFFPIHGVMKITTVPATRGNNRLML 657
Qy 1005 SQVITN--NASC-IVQNLTEPPGPPVHPPEELQHTNRLNGSNTQCQVVRCHLGLQAKGTEV 1061
Db 658 RDPFTDQNTSCNIGWNSYRSTPTE-EDLSHAPQRNHSNDVVSICNL-RLAPSOET 715
Qy 1062 SVGLLRLVHNEFFRRKFKSLTVVSTFELGTGTEGSQLQTEASRWSLSLEVVQTRPTLI 1121
Db 716 SFYLVGNLWLTSLKALKYRSUKITVNAALQRFHSPFIREFDPSRQVTFEISQEDQV 775
Qy 1122 SLWTLIGSVGLGLLLALLVFLCMLKGLFFAHKK-----IPEBEK 1160
Db 776 PIWIVGSLTGLGLLLALLVFLCMLKGLFFAHKK-----IPEBEK 823

RESULT 14
Q8BM12 ID Q8BM12 PRELIMINARY; PRT; 288 AA.
AC Q8BM12;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 6 days neonate skin cDNA, RIKEN full-length enriched
DE library, clone:A030011f23 product:INTEGRIN ALPHA-10 homolog
DE (Fragment).
GN Name=Itgal1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).

```



```
FT REPEAT 456 517 FG-CAP 5.
FT REPEAT 519 577 FG-CAP 6.
FT REPEAT 582 634 FG-CAP 7.
FT CA BIND 467 475 Potential.
FT CA BIND 531 539 Potential.
FT CA BIND 594 602 Potential.
FT SITE 1126 1130 GPKR motif.
FT DISULFID 69 76 By similarity.
FT DISULFID 108 126 By similarity.
FT DISULFID 656 711 By similarity.
FT DISULFID 769 775 By similarity.
FT DISULFID 845 860 By similarity.
FT DISULFID 993 1017 By similarity.
FT DISULFID 1022 1027 By similarity.
FT CARBOHYD 61 61 N-linked (GlcNAc...) (potential).
FT CARBOHYD 89 89 N-linked (GlcNAc...) (potential).
FT CARBOHYD 245 245 N-linked (GlcNAc...) (potential).
FT CARBOHYD 393 393 N-linked (GlcNAc...) (potential).
FT CARBOHYD 696 696 N-linked (GlcNAc...) (potential).
FT CARBOHYD 734 734 N-linked (GlcNAc...) (potential).
FT CARBOHYD 784 784 N-linked (GlcNAc...) (potential).
FT CARBOHYD 907 907 N-linked (GlcNAc...) (potential).
FT CARBOHYD 936 936 N-linked (GlcNAc...) (potential).
FT CARBOHYD 1045 1045 N-linked (GlcNAc...) (potential).
SQ SEQUENCE 1161 AA; 126600 MW; 2258491A984A705E CRC64;

Query Match 18.6%; Score 1136; DB 1; Length 1161;
Best Local Similarity 30.1%; Pred. No. 9e-72;
Matches 363; Conservative 208; Mismatches 486; Indels 148; Gaps 46;

Qy 20 CSPNLDEHRLPPGPEAFGYSVLQHVGGGQRMVLGAPWDPGSDRRGDVYRCPVG 79
Db 17 CHGSLNDEEPIVF-REDAASFGQTVVQF--GSGR-LVVGAPLEAVAVNQTRUYDC--- 69
Qy 80 GAHNAPEAKGHLGQYQLGNSHPVAVNMHLGMSLTDGCGFMACAPLWSRACSSVFS 139
Db 70 ----AP-ATGWCQPIVL-RSPLEAVNMSLGLSLVATNNAQLLACGPTAQCACVKNYAK 123
Qy 140 GICARVDASFQPGSLAPTAQRCTPY-MDVVIVLDGNSI--YPMSEVQTFRLRLVKLFP 196
Db 124 GSCULLGSSLFQIQAVPASMCEPCREQEMDIAFLIDGSGSINQORDFAQMKDFVKALMGE-F 182
Qy 197 IDPQIQVGLVOYGESPVHESLGDFTKEVVRVRAAKNLSRREGRETAKQAINVACTEG 256
Db 183 ASTGTL-FSLMQYSNLIKTHFTFTEFNILDPQSLVDPIVQLGL-TYATGIRTVMEEL 240
Qy 257 FSQSHGGRPEARLLVVVTDGESH-DGEELPAALKAACEAGRVTRYGIAVLGHYLRQRDP 315
Db 241 FHSKNGSRKSAKTLILLVTDQKYRDPLEYSDVIPADKAGIIRYAICVGDAF-----QEP 296
Qy 316 SSFLREIRTIASDPDRFFFNVTDEAALTDIVDALGDRIFGLESHAENESSFGLEMSQI 375
Db 297 TA-LKELNTIGSAPPQDHVFKVGNFAALRSIQRLQEKIFAIEGTQSRSSSSFOHMSQE 355
Qy 376 GFSTHRLKDGILFGMVGAYDGGSVLMLEGGHRLPPP-----RMALEDEFPALQNH 428
Db 356 GFSSALTSDBGVLGAVGFSW-----SGGAFLYPPNTRPTFFINMQEN-----VDMRD 403
Qy 429 AYLGYSSVSMLLRGRRLLFSLCAPFRHRGVIAFOLKKGAVRVAQSLQCEQIGSYFGS 488
Db 404 SYLGYSTAVAFWKGVHSLIL-GARPHQHTGVVIF-TOEARHWRPKSEVRGTQIGSYFGA 461
Qy 489 ELCPLDTRDGTDTVLLVAAPMFLGPQNKETGRVYVILV-CQQSLLLTLQGTLOPEPPQD- 546
Db 462 SLCSVDVDRDGSITDLVLIGAPHYY--EQTRGGQSVFPVPGVRGWQCEATLHGEQGHWP 519
Qy 547 ARFGFAMGALPDNLQDGFADVAVGAPLEDHGOALYLYHG-TQSGVRPHPAQRIAAASMP 605
Db 520 GRFGVALTVLGDVAGNDNLADVAIGAPGEESRGAVIIFHGASRLIEIMPSPQRVTGSQLS 579
Qy 606 HALSYFGRSVDGRLLDGDGDLVDVAVGAQGAAILLSRPIVHLTPSLVPTPAISVWQRD 665
Db 580 LRLQYFGSLSGGQDLTDGGLVDLAVGAQGHVLLRLSLPLKVELSLFAPMEVAKAVYQ 639
```

```
Qy 666 CRRR-----GQEAIVCLTAALCFQVYTSRTFGRWDH-QFYMRFTASLDEWTAGARAADFG 717
Db 640 CWERTPTVLEAGEATVCLT-----VHKGSPDLLGNVQGSVRYDLDALDPGLRLISRAIFDE 693
Qy 718 SQQLSPRRRLSVGNVTCBQLHFHVLVD-TSDYLRPVVALTVTFAL-DNTTKP---GPVLN 772
Db 694 TKNCTLTGRKTLGLGD-HCETVKLLLPDCVEDAVSPIILRLNFSLVSDSASPRNLHPVLA 752
Qy 773 EGSPTSQIKLVFPFSKDCGPNCEVTDLVLQNMDIRSKAPFVVRGGRKRVLVSTTLEN 832
Db 753 VGSQDHITASLPFEKNCKQELLCEGDLGISFNFS-----GLQVLVVGSGSPELTVTVTVN 807
Qy 833 RKNAYNTSLSIIFSRNLHLASLTPORES---PIKVEC-RAPSAAHRL-----CSVGHVPV 884
Db 808 EGSDSYGLTVLKFYYPAGLSVRRVTGTQOPHOYPLRLACEAPAAQEDLRSSCSINHPIF 867
Qy 885 QTGAQVTFLEFBFSCSSLSQVFGKLTASSDSLSERNGLTQENTA-QTSAYIQY----- 937
Db 868 REGAKITFTMITFDVSYKAFGLGDL--LLRAKASENNKPDTNKTAFOLELPVKYTVYTLI 925
Qy 938 -----EPHLLFSS-----ESTLHRYEVHPYGTLPVGPGEFXTTLRVQNLGCVVVSGL 985
Db 926 SRQEDSTNHVNFSSSHGGRQEAHRYRVNLSPL-----KLAVRVNFWFVPLLVNGV 977
Qy 986 IISALLPAVAGGNYFELSQVITNNAASCIVQNLTEPPGPPVHPPELOHTNRLNGSNTQ 1044
Db 978 AVMDVTLLSSPAQG-----VSCVSG--MKPPQNPDPFTQIQRSLVDCSIA 1021
Qy 1045 QVVRCHLGQLAGTEVSVGLLRLVHNEFFRRAKFKSLTVVSTPELGTGEGSVLQL--TE 1102
Db 1022 CLHFRCDIPSLDIQDELDFILRGNLSEGWVSQTLQEKVLLVSEAEITFDTSVYSQLPQE 1081
Qy 1103 A---SRWSESLLEVQTRPILISLWILIGSVLGGILLALLLVFLCWLKGLGFFAHKKIPBE 1159
Db 1082 AFLRAQVETTLSEYVVEPI----FLVAGSVGGLLLALITVVLVKLGFF----- 1128
Qy 1160 KREEK 1164
Db 1129 KRQYK 1133
```

Search completed: April 6, 2005, 12:15:23

Job time : 179.72 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 147.588 Seconds  
(without alignments)  
3058.161 Million cell updates/sec

Title: US-09-647-544-2  
Perfect score: 6106  
Sequence: 1 MEKPFVTHLPFLVFLTGLC.....GFFAHKKIPREKREKLEQ 1167

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 16Dec04:.\*  
1: Geneseqp1980s:.\*  
2: Geneseqp1990s:.\*  
3: Geneseqp2000s:.\*  
4: Geneseqp2001s:.\*  
5: Geneseqp2002s:.\*  
6: Geneseqp2003as:.\*  
7: Geneseqp2003bs:.\*  
8: Geneseqp2004as:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6106	100.0	1167	3	AAY32242 Human int
2	6092	99.8	1167	8	ADQ19290 Human sof
3	6040	98.9	1167	4	AAB64584 Human sec
4	6040	98.9	1167	6	ABP99490 Human sec
5	6040	98.9	1167	6	ABR00964 Human gen
6	6040	98.9	1167	6	ADA4026 Human sec
7	6031	98.8	1152	4	AAB64657 Human sec
8	5965	97.7	1152	4	AAB64658 Human sec
9	5904.5	96.7	1132	3	AAY32243 Human int
10	5080	83.2	1049	5	ADR41424 Human CD-
11	2439.5	40.0	1188	4	AAU14231 Human nov
12	2439.5	40.0	1188	4	ABR50085 Human A25
13	2439.5	40.0	1188	5	AAU10551 Human A25
14	2438.5	39.9	1188	4	AAU14467 Human nov
15	2438.5	39.9	1188	7	ADE09956 Novel pro
16	2429	39.8	1189	3	AAB25582 ITGAl1 pr
17	2429	39.8	1189	4	ABG12949 Novel hum
18	2429	39.8	1189	6	ABR58364 Human NOV
19	2429	39.8	1189	6	ADA27054 Human nov
20	2429	39.8	1189	7	ADE63570 Human Pro
21	2429	39.8	1189	8	ADE86584 Novel hum
22	2429	39.8	1189	8	ADQ19968 Human sof
23	2422	39.7	1188	4	ABR50087 Murine A2
24	2422	39.7	1188	5	AAU10552
25	2403.5	39.4	1188	4	AAB30929 Amino aci

26	2354	38.6	1189	8	ADH80785	Adh80785 Human pol
27	2254	36.9	1034	3	AAB25590	Protein e
28	2254	36.9	1034	6	ADA27062	Human nov
29	2254	36.9	1034	8	ADE86592	Novel hum
30	2254	36.9	1058	5	ADR41496	Human CD-
31	2224.5	36.4	1120	6	ABR58365	Human NOV
32	2037	33.4	987	8	ADP29492	Human sec
33	1894.5	31.0	1180	5	ABR90788	Rat Tumou
34	1894.5	31.0	1180	6	ABU54495	Mouse tum
35	1894.5	31.0	1180	7	ADE83568	Rat Prote
36	1893	31.0	1179	5	ABR90759	Human Tum
37	1893	31.0	1179	5	ABP64915	Human pro
38	1893	31.0	1179	6	ABU54466	Human tum
39	1887	30.9	1151	8	ADE86652	Human int
40	1880	30.8	1177	8	ADQ39498	Human myo
41	1831	30.0	1183	2	AAU07728	Armenian
42	1756	28.8	1179	8	ABO84632	Human can
43	1756	28.8	1181	6	ABU03548	Angiogene
44	1756	28.8	1181	6	ABR59703	Human VLA
45	1756	28.8	1181	6	AAG79775	Alpha2 in

ALIGNMENTS

RESULT 1

AAU32242  
ID AAY32242 standard; protein; 1167 AA.

XX	AC	AAY32242;				
XX	AC	AAY32242;				
DT	15-FEB-2000	(first entry)				
XX	DE	Human integrin subunit alpha-10.				
XX	DE	Human integrin subunit alpha-10.				
KW	KW	Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;				
KW	KW	osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;				
KW	KW	therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker.				
XX	OS	Homo sapiens.				
XX	XX	XX	Location/Qualifiers			
FT	FT	Peptide	1..22	"signal peptide"		
FT	FT	Protein	23..1145			
FT	FT	Domain	23..1120	"mature protein"		
FT	FT	Modified-site	98	"extracellular domain"		
FT	FT	Domain	162..359			
FT	FT	Modified-site	336	"N-glycosylated"		
FT	FT	Modified-site	364	"N-glycosylated"		
FT	FT	Binding-site	494..502	"N-glycosylated"		
FT	FT	Binding-site	558..566	"cation binding site motif"		
FT	FT	Binding-site	620..628	"cation binding site motif"		
FT	FT	Modified-site	733	"cation binding site motif"		
FT	FT	Modified-site	839	"N-glycosylated"		
FT	FT	Modified-site	921	"N-glycosylated"		
FT	FT	Modified-site	1018	"N-glycosylated"		
FT	FT	Modified-site	1039	"N-glycosylated"		
FT	FT	Modified-site	1121..1145	"N-glycosylated"		
FT	FT	Domain				



```

FT FT /note= "transmembrane domain"
FT FT 1122..1167
FT FT /note= "cytoplasmic domain, specifically claimed in Claim
FT FT 21"
XX XX
XX MO9951639-A1.
XX 14-OCT-1999.
XX 31-MAR-1999; 99WO-SE000544.
XX 02-APR-1998; 98SE-00001164.
XX 28-JAN-1999; 99SE-00000319.
XX (ACT1-) ACTIVE BIOTECH AB.
XX Lundgren-Akerlund E;
XX WPI; 2000-052639/04.
XX N-PSDB; AAZ34719.
XX New isolated integrin subunit alpha-10, used as a marker or target
XX molecule for cells during development, regeneration and pathological
XX conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
XX inflammation.
XX Claim 1; Fig 6; 90pp; English.
XX This sequence represents novel human chondrocyte integrin subunit alpha-
XX 10 (Isa10). A splice variant is given in AA3243. The invention relates
XX to a recombinant or isolated integrin heterodimer comprising the alpha10
XX subunit in association with subunit beta (especially beta-1). The
XX heterodimer and the subunit alpha-10 can be used as markers or targets of
XX all types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts.
XX They can also be used: for treating pathological conditions involving
XX Isa10, such as damage to cartilage, trauma, rheumatoid arthritis or
XX osteoarthritis; for detecting the formation of cartilage during embryonal
XX development, physiological or therapeutic reparation of cartilage, or
XX detecting regeneration of cartilage or chondrocytes during
XX transplantation of cartilage or chondrocytes; for selection and analysis
XX or for sorting, isolating or purification of chondrocytes and for in
XX vitro studies of differentiation of chondrocytes; and as a target for
XX anti-adhesive drugs or molecules in tendon, ligament, skeletal muscle or
XX other tissues where adhesion impairs the function of the tissue (all
XX claimed). Isa10 binding entities can be used to determine the
XX differentiation-state of cells during embryonic development, angiogenesis
XX or development of cancer, in pathological conditions such as rheumatoid
XX arthritis, osteoarthritis or cancer, in tissue regeneration or in
XX therapeutic and physiological reparation of cartilage (claimed). A
XX vaccine comprising the integrin heterodimer or subunit alpha-10 is also
XX claimed. Isa10 polynucleotides, vectors, host cells and methods of
XX producing recombinant Isa10 are also claimed
XX
SQ Sequence 1167 AA;
Query Match 100.0%; Score 6106; DB 3; Length 1167;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MELPFVTHLFLPLVFLTGLCSFPLDEHHPLFPFGPEAEFGYSVLQHVGGQRWMLVGA 60
Db 1 MELPFVTHLFLPLVFLTGLCSFPLDEHHPLFPFGPEAEFGYSVLQHVGGQRWMLVGA 60
Qy 61 PWDGSGDRRCVGVGGAHAPCAKGLHGDYQLGNSSHPANMHLGMSLLETDG 120
Db 61 PWDGSGDRRCVGVGGAHAPCAKGLHGDYQLGNSSHPANMHLGMSLLETDG 120
Qy 121 FMACAPLWSRACGSGSVSSGICARVDASFQPGQSLAPTACRCPTVMVIVLDCGNSIYP 180
Db 121 FMACAPLWSRACGSGSVSSGICARVDASFQPGQSLAPTACRCPTVMVIVLDCGNSIYP 180
Qy 181 WSEVOTFLRLVGLFIDPEIQVGLVQGESPVHWSLGDPRFTKEEVVRAAKNLSRREG 240

```

DT 26-AUG-2004 (first entry)  
 DE Human soft tissue sarcoma-upregulated protein - SEQ ID 2109.  
 XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.  
 KW Homo sapiens.  
 OS  
 XX WO2004048938-A2.  
 FN 10-JUN-2004.  
 PD  
 XX 26-NOV-2003; 2003WO-US038193.  
 PF  
 XX 26-NOV-2002; 2002US-0429739P.  
 PR  
 XX (PROT-) PROTEIN DESIGN LABS INC.  
 PA Aziz N, Ginsburg WM, Zlotnik A;  
 XX WPI; 2004-441208/41.  
 DR  
 XX Early detection of soft tissue sarcoma comprises determining expression  
 CC of a gene in a first soft tissue sample and a normal soft tissue sample  
 CC and comparing the gene expression, also useful in treating soft tissue  
 CC sarcoma.  
 PT  
 XX Example 2; SEQ ID NO 2109; 210pp; English.  
 PS  
 XX The invention relates to a novel method for detecting soft tissue sarcoma  
 CC which comprises obtaining a first soft tissue sample from an individual  
 CC and a normal soft tissue sample from the same or different individual,  
 CC determining the expression of a gene in both samples and comparing the  
 CC expression of the gene in both soft tissue samples, where a higher level  
 CC of protein expression in the first soft tissue sample indicates the  
 CC presence of soft tissue sarcoma. The method of the invention has  
 CC cytostatic applications and may be useful for detecting soft tissue  
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
 CC acid sequences may be useful in diagnostic and screening applications.  
 CC The current sequence is that of a human soft tissue sarcoma-upregulated  
 CC protein of the invention. The current sequence is not shown within the  
 CC specification per se but was submitted in CD format by the inventor.  
 CC  
 XX Sequence 1167 AA;  
 SQ  
 Query Match 99.8%; Score 6092; DB 8; Length 1167;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1164; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MELPFVTHLFLPLVFLTGLCSFNNLDEHHPLRFPGPPEAFEGYVLQHVGGQRWMLVGA 60  
 DB 1 MELPFVTHLFLPLVFLTGLCSFNNLDEHHPLRFPGPPEAFEGYVLQHVGGQRWMLVGA 60  
 QY 61 PWDGSGDRGDRVRCVGGGAHNPACAKGHLGDYQLGNSHPAVNMHLMGMSLLTDDGG 120  
 DB 61 PWDGSGDRGDRVRCVGGGAHNPACAKGHLGDYQLGNSHPAVNMHLMGMSLLTDDGG 120  
 QY 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVIVLDGNSNIYP 180  
 DB 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVIVLDGNSNIYP 180  
 QY 181 WSEVOTFLRLRVGLKLFIDPEIQVGLVQYGSPPVHWSLGFRTKEEVVRAAKNLSREG 240  
 DB 181 WSEVOTFLRLRVGLKLFIDPEIQVGLVQYGSPPVHWSLGFRTKEEVVRAAKNLSREG 240  
 QY 241 RETKTAQIMWACTEGFSQSHGRPEAARLLVWTDGSHDGEELPAALKAACEAGRVTY 300  
 DB 241 RETKTAQIMWACTEGFSQSHGRPEAARLLVWTDGSHDGEELPAALKAACEAGRVTY 300  
 QY 301 GIAVLGHLRRQDPSSFLREIRTIASDPDERFFNVTDAAALTDIVDALGDRIFGLEGS 360  
 DB 301 GIAVLGHLRRQDPSSFLREIRTIASDPDERFFNVTDAAALTDIVDALGDRIFGLEGS 360

361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGYDWGCVLWLEGHRLFPFRALEDEF 420  
 DB HAENESSFGLMSQIGFSTHRLKDGILFGMVGYDWGCVLWLEGHRLFPFRALEDEF 420  
 QY 421 PPALQNHAAVLYGYSVSSMLLRGGRRFLSGAPRFRHGKVIATFOLKKGAVRVAAQSLOGE 480  
 DB 421 PPALQNHAAVLYGYSVSSMLLRGGRRFLSGAPRFRHGKVIATFOLKKGAVRVAAQSLOGE 480  
 QY 481 QIGSYFGSELCPDLTDRDGTDLVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLQGTFLQ 540  
 DB 481 QIGSYFGSELCPDLTDRDGTDLVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLQGTFLQ 540  
 QY 541 PEPQDARFGFAMGALPDNLQDGFADVAVGAPLEDHOGALYLVHGTQSGVRPAPQRIA 600  
 DB 541 PEPQDARFGFAMGALPDNLQDGFADVAVGAPLEDHOGALYLVHGTQSGVRPAPQRIA 600  
 QY 601 AASMPHALSYFGRSVDGRLDLDDGDDLDVAVAGAAAILLSSRPVHLTSPLEVTPOAIS 660  
 DB 601 AASMPHALSYFGRSVDGRLDLDDGDDLDVAVAGAAAILLSSRPVHLTSPLEVTPOAIS 660  
 QY 661 VVQRDCRRRGOEAVCLTAALCFQVTSRTPGRWDHOFYMRFTASLDEWTAGARAAFDGSGQ 720  
 DB 661 VVQRDCRRRGOEAVCLTAALCFQVTSRTPGRWDHOFYMRFTASLDEWTAGARAAFDGSGQ 720  
 QY 721 RLSRRLRLSVGNVTCQLHFLHVLDTSDYLRPVALTVTFALDNTTKPGVPLNEGSPTSIQ 780  
 DB 721 RLSRRLRLSVGNVTCQLHFLHVLDTSDYLRPVALTVTFALDNTTKPGVPLNEGSPTSIQ 780  
 QY 781 KLVPFSKDCPDNECVTDLVQNMDIRGSKAPFVVRGGRKVLVSTTLNENKENAYNT 840  
 DB 781 KLVPFSKDCPDNECVTDLVQNMDIRGSKAPFVVRGGRKVLVSTTLNENKENAYNT 840  
 QY 841 SLSIFSRNLHLASLTLPQRESPIKVECAAPSAHARLCSVGHVPFQTGAKVTFLLFEFESC 900  
 DB 841 SLSIFSRNLHLASLTLPQRESPIKVECAAPSAHARLCSVGHVPFQTGAKVTFLLFEFESC 900  
 QY 901 SLLSQVFGKLTASSDSLRNGTLQNTAOTSAYIOEPHLLFSESTLHREYVHPYGTI 960  
 DB 901 SLLSQVFGKLTASSDSLRNGTLQNTAOTSAYIOEPHLLFSESTLHREYVHPYGTI 960  
 QY 961 PVGPGPFKTTLRVQNLGCVVSGLIISALLPAVAGNGYFSLSQVITNNASCIVQNL 1020  
 DB 961 PVGPGPFKTTLRVQNLGCVVSGLIISALLPAVAGNGYFSLSQVITNNASCIVQNL 1020  
 QY 1021 EPPGPPVHPELOHTNRLNGSNTCCQVVRCHLGOLAKGTEVSVGLLRVHNEFFRRAKFK 1080  
 DB 1021 EPPGPPVHPELOHTNRLNGSNTCCQVVRCHLGOLAKGTEVSVGLLRVHNEFFRRAKFK 1080  
 QY 1081 SLTVVSTFELGTEGSLVQLTEASRWSESLLEVVQTRPILISLWILGSLVGLGULLALL 1140  
 DB 1081 SLTVVSTFELGTEGSLVQLTEASRWSESLLEVVQTRPILISLWILGSLVGLGULLALL 1140  
 QY 1141 VFCLWLKLGFFAHKKIPEEKREKLEQ 1167  
 DB 1141 VFCLWLKLGFFAHKKIPEEKREKLEQ 1167  
 RESULT 3  
 AAB64584  
 ID AAB64584 standard; protein; 1167 AA.  
 XX AAB64584;  
 AC  
 XX 22-MAR-2001 (first entry)  
 DT  
 XX Human secreted protein #37.  
 DE  
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.  
 XX







XX OS Homo sapiens.  
 XX PN WO2003000865-A2.  
 XX XX  
 XX PD 03-JAN-2003.  
 XX PF 26-MAR-2002; 2002WO-US009105.  
 XX PR 27-MAR-2001; 2001US-0278650P.  
 XX PR 12-SEP-2001; 2001US-00950082.  
 XX PR 12-SEP-2001; 2001US-00950083.  
 XX XX  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PI  
 XX PI Rosen CA, Ruben SM;  
 XX DR N-PSDB; ADA43832.  
 XX WPI; 2003-184045/18.  
 XX DR  
 XX PT A human secreted protein and nucleic acids useful for preparing a  
 PT diagnostic or pharmaceutical composition for diagnosing or treating  
 PT diabetes or conditions related to diabetes, e.g. hyperglycemia, obesity,  
 PT retinopathy, neuropathy.  
 XX XX  
 XX PS Claim 1; SEQ ID NO 214; 701pp; English.  
 XX CC The invention relates to novel genes and their fragments which are useful  
 CC for preventing, treating or ameliorating medical conditions e.g. by  
 CC protein or gene therapy. The genes are isolated from a range of human  
 CC tissues disclosed in the specification. The nucleic acids and proteins  
 CC are useful in the diagnosis, treatment and prevention of conditions  
 CC related to diabetes, e.g. hyperglycaemia, obesity, retinopathy,  
 CC polynuropathy, atherosclerosis, anaemia, stroke, gangrene, impotence,  
 CC infection, cataract, renal disorders, or endocrine disorders. The present  
 CC sequence was used to illustrate the invention.  
 XX XX  
 XX SQ Sequence 1167 AA;  
 Query Match 98.9%; Score 6040; DB 6; Length 1167;  
 Best Local Similarity 99.0%; Pred. No. 0;  
 Matches 1155; Conservative 3; Mismatches 9; Indels 0; Gaps 0;  
 Qy 1 MELPFVTHLFLPLVFLTGLCSPPNLDHHPRLFPGPPEAEFGYSVLQHVGGQRWMLVGA 60  
 :  
 Db 1 VELPFVTHLFLPLVFLTGLCSPPNLDHHPRLFPGPPEAEFGYSVLQHVGGQRWMLVGA 60  
 Qy 61 PWDGSGDRGDVYRCVPVGGAHNAPCAKGLHLDYQLGNSSHPAVNMHLGMSLLETGDDG 120  
 Db 61 PWDGSGDRGDVYRCVPVGGAHNAPCAKGLHLDYQLGNSSHPAVNMHLGMSLLETGDDG 120  
 Qy 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTMYMDVIVLDGNSNIYP 180  
 Db 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTMYMDVIVLDGNSNIYP 180  
 Qy 181 WSEVOTFLRLRVGLKFLIDPEIQVGLVOYGSPPVHWSLGDFTKKEEVVRAAKNLSRREG 240  
 Db 181 WSEVOTFLRLRVGLKFLIDPEIQVGLVOYGSPPVHWSLGDFTKKEEVVRAAKNLSRREG 240  
 Qy 241 RETYTAQAIWVACTEGSGSQSHGGRPEAARLLVVTGDESHDGEELPAALKAACEAGRVTRY 300  
 Db 241 RETYTAQAIWVACTEGSGSQSHGGRPEAARLLVVTGDESHDGEELPAALKAACEAGRVTRY 300  
 Qy 301 GIAVLGHLRQRDPSSFLRIRTIASDPDERFFFNVTDEAALTDIVDALGDRIFGLEGS 360  
 Db 301 GIAVLGHLRQRDPSSFLRIRTIASDPDERFFFNVTDEAALTDIVDALGDRIFGLEGS 360  
 Qy 361 HAENESSFGLMSQIGFSTHRLKDKILFGMVGAYDWGGSVLWLEGCHRLFPFRMALEDEF 420  
 Db 361 HAENESSFGLMSQIGFSTHRLKDKILFGMVGAYDWGGSVLWLEGCHRLFPFRMALEDEF 420  
 Qy 421 PPALQNHAXYLGYSSXWMLRGXRLXLSGAXRFRHRGKVIAFOLKKGAVVRAQSGLOGE 480

Db 421 PPALQNHAXYLGYSSXWMLRGXRLXLSGAXRFRHRGKVIAFOLKKGAVVRAQSGLOGE 480  
 Qy 481 QIGSYFGSELCPDLTDTRDGTDLVLLVAAPMFLPQNKETGRVYVYLVGQQLLTQGLQ 540  
 Db 481 QIGSYFGSELCPDLTDTRDGTDLVLLVAAPMFLPQNKETGRVYVYLVGQQLLTQGLQ 540  
 Qy 541 PEPQDARFGFAMGALPDNODGFADVAVGAPLEDGHQGALYLHGTQSGVRPHPAQRIA 600  
 Db 541 PEPQDARFGFAMGALPDNODGFADVAVGAPLEDGHQGALYLHGTQSGVRPHPAQRIA 600  
 Qy 601 AASMPHALSYFGRSVDCGRDLDDGDDLDVAVGAGCAAILLSSRPVHLTSPLEVTPOAIS 660  
 Db 601 AASMPHALSYFGRSVDCGRDLDDGDDLDVAVGAGCAAILLSSRPVHLTSPLEVTPOAIS 660  
 Qy 661 VVQRDCRRRGOEAVCLTAALCFQVTSRTPGRDWHQFYMRFTASLDEWTAGARAAFDGSGQ 720  
 Db 661 VVQRDCRRRGOEAVCLTAALCFQVTSRTPGRDWHQFYMRFTASLDEWTAGARAAFDGSGQ 720  
 Qy 721 RLSPRRRLSVGNVTCESQLHFPVLDLSDYLRPVALTVTFALDNTTKPGPVINEGSPISIQ 780  
 Db 721 RLSPRRRLSVGNVTCESQLHFPVLDLSDYLRPVALTVTFALDNTTKPGPVINEGSPISIQ 780  
 Qy 781 KLVPFSDKCGPDNECVTDLVLQVNMDIRGSKAPFVVRGGRKVLVSTTLLENRKENAYNT 840  
 Db 781 KLVPFSDKCGPDNECVTDLVLQVNMDIRGSKAPFVVRGGRKVLVSTTLLENRKENAYNT 840  
 Qy 841 SLSIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPFOTGAKVTFLEFEFSC 900  
 Db 841 SLSIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPFOTGAKVTFLEFEFSC 900  
 Qy 901 SSLSQVFGKLTASSDSLERNGTLQENTQTSAYIOYEPHLLFSESTLHRYEYVHPYCTL 960  
 Db 901 SSLSQVFGKLTASSDSLERNGTLQENTQTSAYIOYEPHLLFSESTLHRYEYVHPYCTL 960  
 Qy 961 PVGPGPEFKTTLRVQNLGCVVSGLIISALLPAVAGHGNFSLSQVTNNASCIVQNL 1020  
 Db 961 PVGPGPEFKTTLRVQNLGCVVSGLIISALLPAVAGHGNFSLSQVTNNASCIVQNL 1020  
 Qy 1021 EPPGPPVHPPELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLRVHNEFFRRAKFK 1080  
 Db 1021 EPPGPPVHPPELQHTNRLNGSNTQCQVVRCHLQGLAKGTEVSVGLLRVHNEFFRRAKFK 1080  
 Qy 1081 SLTVVSTFELGTEGSGVLQLTASRWSESLLEVQVTRPILISLWILGSLGGLLLALL 1140  
 Db 1081 SLTVVSTFELGTEGSGVLQLTASRWSESLLEVQVTRPILISLWILGSLGGLLLALL 1140  
 Qy 1141 VFCLWKLGFPAHKKIPPEEKREKLEQ 1167  
 Db 1141 VFCLWKLGFPAHKKIPPEEKREKLEQ 1167  
 RESULT 7  
 AAB64657  
 ID AAB64657 standard; protein; 1152 AA.  
 XX AAB64657;  
 AC  
 XX  
 DT 22-MAR-2001 (first entry)  
 XX  
 DE Human secreted protein BLAST search protein SEQ ID NO: 167.  
 XX  
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
 KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.  
 XX Homo sapiens.  
 XX OS  
 XX PN WO200077197-A1.  
 XX XX  
 PD 21-DEC-2000.  
 XX











endocrine disorder; reproductive system disorder; infectious disease; gastrointestinal disorder; drug screening; tissue regeneration; chemotaxis; gene therapy; antibody therapy; drug targeting; chromosome mapping; forensic analysis; immunophenotyping; cytostatic; haemostatic; tranquiliser; vulnery; antiinflammatory; nephrotropic; cardiant; antiallergic; anti-HIV; antirheumatic; antiarthritic; antiporiatic; immunosuppressive; vasotrophic; neutropic; neuroprotective; antithyroid; thymimetic; gynaecological; virucide; hepatotropic; antibacterial; dermatological; chromosome 1q21.

OS Homo sapiens.

PN WO200226930-A2.

PD 04-APR-2002.

PF 25-SEP-2001; 2001WO-US029838.

PR 26-SEP-2000; 2000US-0235484P.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Birse CE;

XX WPI; 2002-405050/43.

DR N-PSDB; ADR41248.

Novel polynucleotides and polypeptides useful for treating, preventing or ameliorating cardiovascular, renal, neurovascular, and autoimmune disorders.

PS Claim 11; SEQ ID NO 223; 1243pp; English.

CC The invention relates to 167 novel human CD (cluster of differentiation)-like molecules (ADR41388-ADR41563) and to cDNAs encoding them (seqid:11)-

XX Sequence 1049 AA;

Query Match 83.2%; Score 5080; DB 5; Length 1049;

Best Local Similarity 94.5%; Pred. No. 0;

Matches 990; Conservative 7; Mismatches 31; Indels 20; Gaps 3;

Qy 124 CAPWSRA-CGSSVFSSGICARVDASFOQSLAPTAQRCPTYMDV---VIVLGSNSIY 179  
 Db 18 CLQFWMDCPCGCFIPASG-----KPGTHCPTLPNIHGCHCLGLHLSIY 61  
 Qy 180 PWSEVQTFLLRLVGLKFLTDPEQIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRE 239  
 Db 62 PWSEVQTFLLRLVGLKFLTDPEQIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRE 121  
 Qy 240 GRETKTAQAIWVACTEGFSQSHGRPEARLLVVVTDGESHGDELPALAKACEAGRVTR 299  
 Db 122 GRETKTAQAIWVACTEGFSQSHGRPEARLLVVVTDGESHGDELPALAKACEAGRVTR 181  
 Qy 300 YGIAVLGHLRRQDPSSFLREIRTIASDPDERFFNVTTDBAALTDIVDALGDRIFGLEG 359  
 Db 182 YGIAVLGHLRRQDPSSFLREIRTIASDPDERFFNVTTDBAALTDIVDALGDRIFGLEG 241  
 Qy 360 SHAENESSFGLMSOIGESTHRLKDGILFGVMGAYDWGSLVWLEGGHRLFPFRWALEDE 419  
 Db 242 SHAENESSFGLMSOIGESTHRLKDGILFGVMGAYDWGSLVWLEGGHRLFPFRWALEDE 301  
 Qy 420 FPPALQNHAAVYLGYSVSSMLRGGRRFLSGAPFRHRGKVIAFOLKKGAVRVAQSLQG 479  
 Db 302 FPPALQNHAAVYLGYSVSSMLRGGRRFLSGAPFRHRGKVIAFOLKKGAVRVAQSLQG 361  
 Qy 480 EOIGSYFSGELCPDITDRDGTDLVLLVAAAPFLGPNKQKTRGVYVYLVGQSSLLTLQGT 539  
 Db 362 EOIGSYFSGELCPDITDRDGTDLVLLVAAAPFLGPNKQKTRGVYVYLVGQSSLLTLQGT 421  
 Qy 540 QPEPPQDARFGFAMGALPDNODGADVAVGAPLEDHOGALYLYHGTQSGVRPHPAORI 599  
 Db 422 QPEPPQDARFGFAMGALPDNODGADVAVGAPLEDHOGALYLYHGTQSGVRPHPAORI 481

Qy 600 AAASMPHALSYFGRSVDRGLDLGGDDLVVAVAGQAAILSSRPVHLTPSLEVTPOAI 659  
 Db 482 AAASMPHALSYFGRSVDRGLDLGGDDLVVAVAGQAAILSSRPVHLTPSLEVTPOAI 541  
 Qy 660 SVVQDRCRRRGOEAVCLTAALCFQVTSRTQGRWDHQFVFRFTASLDEWTAGARAADGSG 719  
 Db 542 SVVQDRCRRRGOEAVCLTAALCFQVTSRTQGRWDHQFVFRFTASLDEWTAGARAADGSG 601  
 Qy 720 QRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALVTTFALDNTTKPGPVINEGSPTSI 779  
 Db 602 QRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALVTTFALDNTTKPGPVINEGSPTSI 661  
 Qy 780 QKLVPPFSKDCGPDNECVTDLVQVNMDIRGSRKAPFVVRGRRKVLVSTTLENKENAYN 839  
 Db 662 QKLVPPFSKDCGPDNECVTDLVQVNMDIRGSRKAPFVVRGRRKVLVSTTLENKENAYN 721  
 Qy 840 TSLSIIFSRNLHLASLTPQRESPIKVECAAPSAHARLCSVGHVPVFTQGAKTFLLEPEFS 899  
 Db 722 TSLSIIFSRNLHLASLTPQRESPIKVECAAPSAHARLCSVGHVPVFTQGAKTFLLEPEFS 781  
 Qy 900 CSSLLSQVFGKLTASSDSLERNGTLQENTAQTSAYIQVEPHLLFSSESTLHRYEVHPYGT 959  
 Db 782 CSSLLSQVFGKLTASSDSLERNGTLQENTAQTSAYIQVEPHLLFSSESTLHRYEVHPYGT 841  
 Qy 960 LPVGPGEFETTLRVQNLGCVVSGLLISALLPAVAGGNYFLSLSQVITNASCIVQNL 1019  
 Db 842 LPVGPGEFETTLRVQNLGCVVSGLLISALLPAVAGGNYFLSLSQVITNASCIVQNL 901  
 Qy 1020 TEPGPPVHPPEELOHTNRLNGSNQCCQVVRCHLGQAKGTEVSGLLRVLVNEPFRRAKF 1079  
 Db 902 TEPGPPVHPPEELOHTNRLNGSNQCCQVVRCHLGQAKGTEVSGLLRVLVNEPFRRAKF 961  
 Qy 1080 KSLTVVSTFELGTBEGSVLQLTEASRWSESLEVVQTRPILISLWILIGSVLGGLLLAL 1139  
 Db 962 KSLTVVSTFELGTBEGSVLQLTEASRWSESLEVVQTRPILISLWILIGSVLGGLLLAL 1021  
 Qy 1140 LVFCLWKLGFPAHKKIPEEEKREKLEQ 1167  
 Db 1022 LVFCLWKLGFPAHKKIPEEEKREKLEQ 1049

RESULT 11

AAU14231

ID AAU14231 standard; protein; 1188 AA.

XX AC AAU14231;

XX DT 24-OCT-2001 (first entry)

XX DE Human novel protein #102.

XX KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;

XX KW immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;

XX KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;

XX KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;

XX KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;

XX KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;

XX KW tissue regeneration; immune disorder.

XX OS Homo sapiens.

XX PN WO200155437-A2.

XX PD 02-AUG-2001.

XX PP 25-JAN-2001; 2001WO-US002623.

XX PR 25-JAN-2000; 2000US-00491404.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;



FT	Domain	/label= Integrin_alphasubunit_repeat_domain_#1
FT	FT	115..157
FT	Domain	/label= Integrin_alphasubunit_repeat_domain_#2
FT	FT	164..345
FT	Domain	/label= I domain
FT	FT	367..392
FT	Domain	/label= Integrin_alphasubunit_repeat_domain_#3
FT	FT	421..455
FT	Domain	/label= Integrin_alphasubunit_repeat_domain_#4
FT	FT	478..516
FT	Domain	/label= Integrin_alphasubunit_repeat_domain_#5
FT	FT	540..575
FT	Domain	/label= Integrin_alphasubunit_repeat_domain_#6
FT	FT	602..640
FT	Domain	/label= Integrin_alphasubunit_repeat_domain_#7
FT	FT	1142..1164
FT	Domain	/label= Transmembrane_domain
FT	FT	1185..1198
FT	Domain	/label= Cytoplasmic_domain
XX	XX	
PN	WO200073339-A1.	
XX	XX	
PD	07-DEC-2000.	
XX	XX	
PF	15-MAY-2000; 2000WO-US013262.	
XX	XX	
PR	28-MAY-1999; 99US-00322790.	
PR	27-APR-2000; 2000US-00561363.	
XX	XX	
PA	(WILL-) MILLENNIUM PHARM INC.	
XX	XX	
PI	Pan Y, Lora JM;	
XX	XX	
DR	WPI: 2001-041142/05.	
DR	N-PSDB; AAC91901, AAC91902.	
XX	XX	
PT	Nucleic acid encoding alpha-integrin subunits, useful for treatment and	
PT	diagnosis of fibrosis, e.g. of the liver.	
XX	XX	
PS	Claim 8; Fig 1; 164pp; English.	
XX	XX	
CC	The present sequence is human integrin alpha subunit, A259. A259 is	
CC	homologous with the alpha1 and alpha10 integrin subunits and is	
CC	overexpressed in fibrosis. A259 is implicated in regulation of	
CC	proliferation, differentiation and/or function of many different cell	
CC	types. Inhibitors of A259 activity are useful for the treatment of liver	
CC	disease, particularly fibrosis, and also fibrosis in other organs	
CC	(specifically lung and kidney). In addition, A259 can be used for	
CC	treatment and prevention of cancer, osteoporosis, acute myeloid	
CC	leukaemia, HIV infection, and rheumatoid arthritis	
XX	XX	
SQ	Sequence 1188 AA;	
	Query Match	40.0%; Score 2439.5; DB 4; Length 1188;
	Best Local Similarity	43.2%; Pred. No. 8.5e-208;
	Matches 517; Conservative 213; Mismatches 421; Indels 45; Gaps 16;	
Qy	1 MELPFVTHLFLPIVFLTCLCSFFNLDEHHPLRFPQPEAEFGYSVLQHVGGQRMVLVGA	60
Db	1 MDLPRGLVWAWALSLWPGFTDTFNMNDRKPRVIFGSRTAFFGYTQQQHDISGNKWL VVGA	60
Qy	61 PWDGSGDRRGDVYRCPVGGAAHAPCAKGHLDGYQLGNSHPAVNMHLGMSLLETDDGG	120
Db	61 PLENTNGYQKTDVYKCPV---IHGNCYKLNLRVTLNVNSERKNMRGLGLSLATNPKDNS	117
Qy	121 FMACAPLWSRACGSSVFSSGICARVDASFPQGS LAPTAQRCPYMDVYVILDGSSNIYP	180
Db	118 FLACSPLWSHECGSSYYTTGCSRWSNFRFSKTVAPALQRCQYTMDIVILDGSSNIYP	177
Qy	181 WSEVQTFLRLVGLKFTIDPEQIQVGLVOYGSPVHEWSLGFRTKEEVVRAAKNLSRREG	240
Db	178 WVEYQHFLLINLKFYTGPGQIQGVVQYGDGVVHFHFLNDRYSVKDVVEAASHIEQRGG	237

Human A259 polypeptide.

Human, A259; integrin alpha subunit; integrin alpha 10; secreted protein; liver disease; fibrosis; lung; kidney; bone associated disorder; blood; cartilage associated disorder; hematopoietic disorder; bone marrow; immune related disease; apoptotic disorder; neuronal tissue disease; neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic; nephrotropic; immunomodulator; anti-inflammatory; neuroprotective; antiarthritic; antianemic; antiallergic; antiasthmatic; dermatological; antidiabetic; anticonvulsant; antiparkinsonian.

Homo sapiens.

Key Location/Qualifiers  
Domain 1..1141  
/note= "Extracellular domain"  
Peptide 1..22  
/note= "Signal peptide"  
Protein 23..1188  
/note= "Mature human A259"  
Domain 37..90  
/note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"  
Domain 115..157  
/note= "Integrin alpha repeat domain"  
Domain 164..345  
/note= "I domain or Von Willebrand Factor type A domain"  
Domain 367..392  
/note= "Integrin alpha repeat domain"  
Domain 421..472  
/note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"  
Domain 476..532  
/note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"  
Domain 538..593  
/note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"  
Domain 600..654  
/note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"  
Domain 1142..1164  
/note= "Transmembrane domain"  
Domain 1165..1188  
/note= "Cytoplasmic domain"

WO200181414-A2.

01-NOV-2001.

27-APR-2001; 2001WO-US013516.

27-APR-2000; 2000US-00561263.

(MILL-) MILLENNIUM PHARM INC.

Pen Y, Lora J;

WPI; 2002-041397/05.

N-PSDB; AAS16873.

New A259 nucleic acids and polypeptides, which comprise integrin alpha subunit, useful for diagnosing, preventing or treating e.g. liver disease, kidney or lung fibrosis, cancers, blood disorders or immune related diseases.

Claim 9; Fig 1; 168pp; English.

The invention relates to human and murine A259 nucleic acid molecules which encode secreted proteins with homology to integrin alpha subunits, specifically to integrin alpha 10. The A259 polypeptide and nucleic acid are useful for treating liver disease or fibrosis, particularly kidney fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also

CC useful for diagnosing, preventing or treating cartilage and bone associated disorders (such as bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and osteoporosis), bone marrow, blood and hematopoietic disorders (such as acute myeloid leukemia, haemophilia, anaemia and thalassemia), immune related diseases (such as HIV, viral infections, cancers, T cell autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g. asthma and psoriasis), apoptotic disorders (such as systemic lupus erythematosus and insulin-dependent diabetes mellitus), diseases of the neuronal tissues (such as epilepsy and muscular dystrophy) and neurodegenerative diseases (such as Parkinson's disease and Huntington's disease). This sequence represents the human A259 polypeptide

XX SQ Sequence 1188 AA;

Query Match 40.0%; Score 2439.5; DB 5; Length 1188;  
Best Local Similarity 43.2%; Pred. No. 8.5e-208; Indels 45; Gaps 16;  
Matches 517; Conservative 213; Mismatches 421;  
Qy 1 MELPFVTHLEPLVFLTGLCSFNLDEHPRLPFGPPPEARFYSVLQHVGGGQWMLVGA 60  
Db 1 MDLPRGLVVAWALSMLWPGFTDTFNMTRKPRVIPGSRTPAFGYTVQQHDISGNKMLVGA 60  
Qy 61 PMDGSGDRRGDYRCVPGGAHNAPCAKHGLDYQLGNSHHPAVNMHLGSLLETDDGG 120  
Db 61 PLETNGYQKTGDVYKCPV---IHGNCCTKLNLRVTLSNVSEKDNMRLGLSLATNPXDNS 117  
Qy 121 PMACAPLWSRACSSVPSFGICARVDASFOQGS LAPTAQRCPTMDVTVLQDGSNIYP 180  
Db 118 FLACSPWSHECSSYTTQMC SRVNSFRSKTVAPALQRCOTYMDIVVLQDGSNIYP 177  
Qy 181 WSEVQTELRRLVGLKFLIDPEQIQVLQYGESVHEWLSLGDPRTKERVVAANKLSREG 240  
Db 178 WVEVQHEPLINLLKFFIGPQIQGVVQYGEDVVEHFLNDYRSKDVVEAASHIEQGG 237  
Qy 241 RETKTAQIMVACTEGSPQSHGGRPEARLLVVVTGESHGDELPAALKACAGRVTRY 300  
Db 238 TETRTAFGIFARSEAFQK--GGRKGAKKVMIVITDGESHSDPLEKVIQQSRDNTVRY 295  
Qy 301 GIAVLGHVLRQRDPSSFLREIRTIASDPDEREFENVTDRAALTDIVDALGDRIFGLEGS 360  
Db 296 AVAVLGYNNRRGINPETFLNEIKYIASDDPDQDPFNVTDRAALKDIDVALGDRIFSLGEGT 355  
Qy 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGYDMGGSVLMLEGGHRLPPPRMALBDEF 420  
Db 356 N-KNETSFGLMSQTGFSSHVEDVLLGAVGAYDMGAVLKETSACKVPLRESYLKEP 414  
Qy 421 PPALQNHAAVLYGSVSMSLLRGGRRRLPLSGAPRHRHGKVIAFQKKDGAVRVAQSLQGE 480  
Db 415 PEELKNHGAYLGVTTSVSSRQGRVYVAGAPRPNHTGKVLFTMNNRSLTIHQAMRGQ 474  
Qy 481 QIGSYFGSELCPDTRDGTDLVLLVAAPMLPGPQNKETGRVYVLYVQQSLLTLOQTLLQ 540  
Db 475 QIGSYFGSEITSDVIDDGDGVTDLVVGAPMYFN--EGRERKVVYVEL--RONRFVYNGTLK 532  
Qy 541 PEPP-QDARFGFAMGALFDLNLQDGFADVAVGAPLEDHQGALYLYHGQTQSGVRPHPAQRI 599  
Db 533 DSHSYQNARFGSSIASVRDLNQDSYNDVVVGAPLEDNHAGAIYIFHGRGSILKTPQRI 592  
Qy 600 AASMPHALSYFGSRVDGRDLDDGDDLDVAVAGQGAAILLSRPPIVHLTPSLVTPQAI 659  
Db 593 TASELATGLQYFGCSIHGQLDLNEDGLIDLAVGALGNVILWSRPVYVQINASLHFEPSKI 652  
Qy 660 SVVORDCRRRGOEAVCLTAALCQVTSRTGRWDHQFYMEFTASLDTSWAGARAAPFGSG 719  
Db 653 NIFHRDCKRGRDATCLAAFLCPTPIFLAHPFQTTVIGIRYNATMDERRRTPRAHLDEGG 712  
Qy 720 QRLSPRLRLSVGNVTCQLHFHFLVLDTSYLRLPVALTVTFALDNTTKPGPVNLGSPSTSI 779  
Db 713 DRFTNRAVLSSGQELCERINFHVLDTADYVKPVPFSVEYSLEDDPH-GPMLDGMWPTTL 771  
Qy 780 QKLVPFSKDCGPDNECVTDVLQVNMDI-----RGRK-----APFVV 817



Db 772 RVSPFWNGCDEHCVPLDLVLDARSDLPTAMEYQORVLRKPAQDCSAYTLSPDTTVFII 831  
 Qy 818 RGRKRVKLVSTLENRKENAYNTLSIIFRNHLASLTPORESPIKVECAAPS--AAH 875  
 Db 832 ESTQRVAVATELNENGENAYSTVLNISQSANLQFASLIQKEDSDGIECVNERRLQKQ 891  
 Qy 876 LCSVGHPVFTQAKVTPLLEFEFSCSLLSQVFGKLTASSDSLERNCTLOENTAQTSAI 935  
 Db 892 VCNVSYFFRAKAVAPRLDPEFSKIFLHHLELELAAGSDSNERDSTKEDNVAFLRHL 951  
 Qy 936 QYEPHLLFSSESTLHYEVHPYGTLP--VGPGEFKTTLRVQNLGCVVSGLIISALLPA 993  
 Db 952 KYEADVLPTRSSLSHVEKPNSSLYRVDGIGPPFCIFRIQNLGLPIHGMWKTIPI 1011  
 Qy 994 VAGGNVFLSLQVITN--NASC-IVQNLTPEPPGVPHPELOHNLNLSNNTQCQVRC 1050  
 Db 1012 ATRSGNRLKLRDPLTDEANTSCNMGNSTEYRTPVE--EDLRRAPQNLHNSDVVSINC 1070  
 Qy 1051 HLGOLAGTEVSGLLRVHNEPRRAKFKSLTVVSTFELGTBEGSVLQLTEASRWESL 1110  
 Db 1071 NI-RLVNPQENFHLGNLWLRSLKALKYKSMKIMVNAALQRPQHPSPFIFREEDPSQIV 1129  
 Qy 1111 LEVVQTRPILISLWILGSLVGLGILLALLVFCILWKLGFPAHKKIPBEEKLE 1166  
 Db 1130 FEISKQEDWQPIWIVGSTLGLLALLVALLVWLKLGFFRSAR---RRREPCLD 1181  
 RESULT 14  
 AAU14467  
 ID AAU14467 standard; protein; 1188 AA.  
 AC AAU14467;  
 XX  
 XX 24-OCT-2001 (first entry)  
 DE Human novel protein #338.  
 XX  
 KW Human; novel protein; anti-naemic; osteopathic; anti-inflammatory;  
 KW immunomodulatory; cytostatic; neuroprotective; vulnary; nootropic;  
 KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;  
 KW antibacterial; antiallergic; dermatological; haemostatic; antiaesthetic;  
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
 KW tissue regeneration; immune disorder.  
 XX  
 OS Homo sapiens.  
 PN W0200155437-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 XX 25-JAN-2001; 2001WO-US002623.  
 XX  
 XX 25-JAN-2000; 2000US-00491404.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Drmanac RT;  
 XX  
 XX WPI: 2001-451939/48.  
 XX  
 XX N-PSDB; AAS22772.  
 XX  
 XX Isolated polypeptides useful for treating anti-inflammatory diseases,  
 XX nervous system disorders, and for regenerating bone and cartilage.  
 XX  
 XX Example 4; Page 828-831; 894pp; English.  
 XX  
 CC The invention relates to polynucleotides encoding novel human proteins or  
 CC their active domains. The polypeptides, polynucleotides and antibodies  
 CC raised against the polypeptides are used in a method of treatment of a  
 CC mammal and prevention of disorders caused by the aberrant protein  
 CC expression or activity. The polypeptides can be used as molecular weight  
 CC markers, food supplements, and in antibody production. The polypeptides

CC are used to identify compounds which bind to the polypeptides.  
 CC Polynucleotides of the invention are used as probes and primers, for  
 CC sequencing, for chromosome or gene mapping, in the production of  
 CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene  
 CC therapy. Polypeptides of the invention can be used to target drugs to a  
 CC tumour, in assays to determine biological activity, to raise  
 CC antibodies/elicite an immune response, to determine quantitative protein  
 CC levels, as tissue markers, and to isolate receptors or ligands.  
 CC Polypeptides of the invention may also be useful in treating platelet  
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
 CC the proliferation, differentiation and survival of stem cells, as a  
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-  
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory  
 CC diseases, nervous system disorders, and infection. The present sequence  
 CC represents a protein of the invention  
 XX  
 XX Sequence 1188 AA;  
 SQ  
 Query Match 39.9%; Score 2438.5; DB 4; Length 1188;  
 Best Local Similarity 43.2%; Pred. No. 1e-207;  
 Matches 517; Conservative 213; Mismatches 421; Indels 45; Gaps 16;  
 Qy 1 MELPFTVTHLPFLPVFTLGLCSFNLDBHPRLPFGPPEAFEGYSVLQHVGGQGRWMLVGA 60  
 Db 1 MDLPGRGLVAVAWALSLMPFGFTDFNMDTRKPRVPGSRTAFPGYTVQOHDISGNKWLAVGA 60  
 Qy 61 PWDGPGSDRGDYYRCPVPGCAHNAKCAKHLGDLQYQNGSSHPAVNMHLGSLLETDDGG 120  
 Db 61 PLETNQYQKTDGVYKCPV---IHGNCNKLNLGRVTLNSVSRKDNMRGLSLATNPKONS 117  
 Qy 121 FMACAPLWSRACGSSVFSFGICARVDASFOQSGSLAPTAQRCPTYMDVIVLDGNSIYP 180  
 Db 118 FLACSLWSHECGSSYYTTCWCSRVNSNFRFSKTVAPALQRCQTYMDIVLDGNSIYP 177  
 Qy 181 WSEVQTLRLRLVKLFIDPEIQVGLVQYGESPVHWSLGDFTKEEVRAAKNLSRREG 240  
 Db 178 WVEVQHFLINILKFKYIGPQIQVGVVQYGEDVHFEHLNDRSVKDVAAASHIEQRGG 237  
 Qy 241 RETKTAQAINVACTEGFSQSHGGRPEARLLVVVTDGESHGDELPALAKACEAGRVTRY 300  
 Db 238 TETRTAFGIEFARSEAFQK--GGRKGAKKVMIVITDGESHSDSPLEKVIQOOSRDNTRY 295  
 Qy 301 GIAVLGHYLRQRDPSSFLREIRTIASDPDERFFNVNTDEAALTDIVDALGDRIFGLEGS 360  
 Db 296 AVAVLGYNNRRGINPEIFLNEIKYIASDPDKIFFNVNTDEAALKDIYDALGDRIFSLEGT 355  
 Qy 361 HAENESSFGLEMSQIGFSTHRLKDGILFGMVGAYDWGSGVLMLEGHRLFPFRMALEDEF 420  
 Db 356 N-KNETSFGLMSQIGFSSHWVEDGVLGAVGAYDMNGAVLKETSACKVPLRESYLKEF 414  
 Qy 421 PPALONHAAVLYGYSVSSMLRGGRRLLFSLGAPRHRGKVIATOLKKGAVRVAOSLQGE 480  
 Db 415 PEELKNHGAYLGYTVTSVSSRQGRVYVAGAPRNFHTGKVLFTMHNRRSLTIHQAMRGQ 474  
 Qy 481 QIGSYFSGELCPDTRDGTDLVLLVAPMFLGPQNKETGRVYVYLVGQGSLLTLQGTLO 540  
 Db 475 QIGSYFSGEITSVDIDGDTVDLLVAGAPMYFN--EGREKGVYVEL--RQNFVYNGTLK 532  
 Qy 541 PEPP--QDAREFGAMGALPDNLQDGFADVAVGAPLEDGHQGLALYLYHGTQSGVRPHPAQRI 599  
 Db 533 DSHSYQNARFGSSIASVRDLNQDSYNDVVVGAPELDNHAGAIYIFHGRGSIKTPQRI 592  
 Qy 600 AAASMPHALSYFGESVDGRDLDDGLDVLVDVAVAGQGAAILLSRPPIVHLTPSLEVTQAI 659  
 Db 593 TASELATGLQYFGCSIHGOLDLNEGLDLAVGALGNVILWSRPVQVQINASLHFEPSKI 652  
 Qy 660 SVVQDCRRRGOEAVCLITAAALCFQVTSRTSGRWDHQFVYMRFTASLDSEWTAGARAAFGSG 719  
 Db 653 NIFHRDCKRSGRDATCLAAFLCTPIFLAFHFQTTVIGIRYNATWDERRYTPRAHLEGG 712





Db 772 RVSVFPWNGCNEDEHCVFDLVLDDARSDLPTAMEYQORVLRPAQDCSAYTSLFDTTTFII 831  
Qy 818 RGRKVLVSTTLNRKENAYNTSLIIFSRNLHLASLTPORESPIKVECAAPS--AHAR 875  
Db 832 ESTRQVAVEATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSGIECVNEERRLQKQ 891  
Qy 876 LCSVGHVFQTAGAKVTFLLPEPESCSSLISQVFGKLTASSDSLERNGLQENTAQTSAYI 935  
Db 892 VCNVSYPPFRAKAKVAFRLDFEFKSIPLHHLEIELAAGSDSNERDSTKEDNVAPLRFHL 951  
Qy 936 QYEPHLLFSSESTLHRYEVHPYGTLP--VGPGEFKTTLRVQNLGCYVWSGLIISALLPA 993  
Db 952 KYEVDVLFTRSSSLSHYEVKPNSSLERYDVGIGPPFCIFRIQNLGLFPIHGMWKITPI 1011  
Qy 994 VAHGGNYFLSLSQVITN--NASC-IVQNLTEPPGPPVHPPELOHTNRNGSNTOCQVVR 1050  
Db 1012 ATRSGNRLKLRLDFTLDEANTSCNIWGNSTYRPTVE-EDLRRAPQLNHSNSDWSINC 1070  
Qy 1051 HLGOLAKGTEVSVGLRLVHNEFFERRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESL 1110  
Db 1071 NI-RLVPNQEIFHLLGNLRLSKALKYKSMKIMVNAALQORQFHPFIFREEDPSRQIV 1129  
Qy 1111 LEVVQTRPILISLWILIGSVLGGLLALLVFCWLKLGFFAHKKIPEEEKREEKLE 1166  
Db 1130 FEISKQEDWQVPIWIIVGSTLGGLLALLVALLWKLGGFFRSAR----RRREPGLD 1181

Search completed: April 6, 2005, 12:22:27  
Job time : 154.588 secs

*The [illegible]*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:17:21 ; Search time 145.761 Seconds  
(without alignments)  
2658.060 Million cell updates/sec

Title: US-09-647-544-2  
Perfect score: 6106  
Sequence: 1 MELPFVTHLFLPLVLTGLC.....GFAHKKIPBEKREKLEQ 1167

Scoring table: BLOSOM62  
Gapop 10.0 , Gapext 0.5

Searched: 1418010 seqs, 331997259 residues

Total number of hits satisfying chosen parameters: 1418010

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep.\*

2: /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB.pep.\*

3: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*

4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*

5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep.\*

6: /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pep.\*

7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep.\*

8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep.\*

9: /cgn2\_6/ptodata/1/pubaa/US09A\_PUBCOMB.pep.\*

10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pep.\*

11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pep.\*

12: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pep.\*

13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pep.\*

14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep.\*

15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep.\*

16: /cgn2\_6/ptodata/1/pubaa/US10D\_PUBCOMB.pep.\*

17: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pep.\*

18: /cgn2\_6/ptodata/1/pubaa/US11\_NEW\_PUB.pep.\*

19: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep.\*

20: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6092	99.8	1167	US-10-741-601-531	Sequence 531, App
2	6092	99.8	1177	US-10-741-601-532	Sequence 532, App
3	2439.5	40.0	1188	US-10-291-265-338	Sequence 338, App
4	2438.5	39.9	1188	US-10-291-265-338	Sequence 810, App
5	2429	39.8	1189	US-09-984-130-35	Sequence 35, Appl
6	2429	39.8	1189	US-09-984-130-35	Sequence 35, Appl
7	2429	39.8	1189	US-10-262-839-43	Sequence 4, Appl
8	2254	36.9	1034	US-09-984-130-43	Sequence 43, Appl
9	2254	36.9	1034	US-09-836-353A-43	Sequence 43, Appl
10	2244.5	36.4	1120	US-10-262-839-6	Sequence 6, Appl
11	1894.5	31.0	1180	US-09-918-715-307	Sequence 307, App
12	1893	31.0	1179	US-09-918-715-250	Sequence 250, App
13	1887	30.9	1151	US-09-984-130-103	Sequence 103, App

14	1887	30.9	1151	10	US-09-836-353A-103	Sequence 103, App
15	1880	30.8	1177	17	US-10-741-600-1161	Sequence 1161, App
16	1756	28.8	1181	14	US-10-160-354-2	Sequence 2, Appl
17	1756	28.8	1181	15	US-10-295-027-1286	Sequence 1286, App
18	1756	28.8	1181	15	US-10-211-462-187	Sequence 187, App
19	1740.5	28.5	1148	16	US-10-872-198-147	Sequence 147, App
20	1727.5	28.3	707	9	US-09-764-870-313	Sequence 313, App
21	1727.5	28.3	707	14	US-10-125-540-313	Sequence 313, App
22	1720	28.2	1147	15	US-10-336-603A-42	Sequence 42, Appl
23	1131.5	18.5	1161	9	US-09-350-259-2	Sequence 2, Appl
24	1131.5	18.5	1161	9	US-09-891-943-2	Sequence 99, Appl
25	1126	18.4	1161	9	US-09-350-259-99	Sequence 99, Appl
26	1126	18.4	1161	9	US-09-891-943-99	Sequence 55, Appl
27	1124.5	18.4	1161	9	US-09-350-259-55	Sequence 55, Appl
28	1124.5	18.4	1161	10	US-09-891-943-55	Sequence 53, Appl
29	1109	18.2	1161	9	US-09-350-259-53	Sequence 53, Appl
30	1109	18.2	1161	10	US-09-891-943-53	Sequence 37, Appl
31	1108.5	18.2	1151	9	US-09-350-259-37	Sequence 37, Appl
32	1108.5	18.2	1151	10	US-09-891-943-37	Sequence 1088, App
33	1106	18.1	1170	17	US-10-741-600-1088	Sequence 2, Appl
34	1105	18.1	1170	15	US-09-945-265-2	Sequence 1, Appl
35	1105	18.1	1170	15	US-10-261-164-1	Sequence 624, App
36	1104.5	18.1	688	10	US-09-866-050A-624	Sequence 295, App
37	1103.5	18.1	1223	16	US-10-408-765A-295	Sequence 1086, App
38	1103.5	18.1	1223	17	US-10-741-600-1086	Sequence 46, Appl
39	1103	18.1	1155	9	US-09-350-259-46	Sequence 46, Appl
40	1103	18.1	1155	10	US-09-891-943-46	Sequence 130, App
41	1100.5	18.0	1145	16	US-10-872-198-130	Sequence 3, Appl
42	1085.5	17.8	1153	9	US-09-350-259-3	Sequence 1, Appl
43	1085.5	17.8	1153	10	US-09-902-481A-1	Sequence 3, Appl
44	1085.5	17.8	1153	10	US-09-891-943-3	Sequence 30, Appl
45	1085.5	17.8	1153	14	US-10-144-259-30	

ALIGNMENTS

RESULT 1

US-10-741-601-531

; Sequence 531, Application US/10741601

; Publication No. US20040166519A1

; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele et al.

; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

; FILE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CLO01500

; CURRENT APPLICATION NUMBER: US/10/741.601

; CURRENT FILING DATE: 2003-12-22

; NUMBER OF SEQ ID NOS: 26415

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 531

; LENGTH: 1167

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-741-601-531

Query Match 99.8%; Score 6092; DB 16; Length 1167;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1164; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MELPFVTHLFLPLVLTGLCSPFNLDHHPRLFGPPEAEFGYSVLQHVGGQRMWLYGA	60
Db	1	MELPFVTHLFLPLVLTGLCSPFNLDHHPRLFGPPEAEFGYSVLQHVGGQRMWLYGA	60
Qy	61	PWDGSGDRRGDVYRCVPVGGAHNAPCAKGLDYLQGNSSHPAVNMHLMGMSLLETGDDGG	120
Db	61	PWDGSGDRRGDVYRCVPVGGAHNAPCAKGLDYLQGNSSHPAVNMHLMGMSLLETGDDGG	120
Qy	121	FMACAPLWSRACGSSVSSGICARVDASFPQGSGLAPTAQRCPTMYDVIIVLDGNSNIYP	180
Db	121	FMACAPLWSRACGSSVSSGICARVDASFPQGSGLAPTAQRCPTMYDVIIVLDGNSNIYP	180
Qy	181	WSEVQTFELRLVGLKFLDIDPEQIQVLQVQSGSPVHWSLGFRTKEEVVRAAKNLSRREG	240

```
Db 181 WSEVQTLRLRLVGLFIDPEIQVGLVQGESPVHWSLGDFTKEEVVRAAKNLSRREG 240
Qy 241 RETKTAQAIWVACTEGFSQSHGGPEARLLVVVTDGSHDGBELPAALKACAGRVTRY 300
Db 241 RETKTAQAIWVACTEGFSQSHGGPEARLLVVVTDGSHDGBELPAALKACAGRVTRY 300
Qy 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDRPFNFVNTDEAALTDIVDALGDRIFGLEGS 360
Db 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDRPFNFVNTDEAALTDIVDALGDRIFGLEGS 360
Qy 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGYDVGSGVLMLEGGHRLFPFRMALEDEF 420
Db 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGYDVGSGVLMLEGGHRLFPFRMALEDEF 420
Qy 421 PPALQNHAAVLYGYSVSMLLRGGRRLFLSGAPRFRHGKVIAPQLKKGAVRVAQSLQGE 480
Db 421 PPALQNHAAVLYGYSVSMLLRGGRRLFLSGAPRFRHGKVIAPQLKKGAVRVAQSLQGE 480
Qy 481 QIGSYFGSELCPDLTDGDTTDLVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLOQTLO 540
Db 481 QIGSYFGSELCPDLTDGDTTDLVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLOQTLO 540
Qy 541 PEPPQDARFGFAMGALPDNLQDGFADVAVGAPLEDHGQALYLYHGTSQSVRPHPAQRIA 600
Db 541 PEPPQDARFGFAMGALPDNLQDGFADVAVGAPLEDHGQALYLYHGTSQSVRPHPAQRIA 600
Qy 601 AASMPHALSYFGRSVGRDLDDGDDLDVDAVGAAGAAILLSSRPVHLTPSLEVTPOAIS 660
Db 601 AASMPHALSYFGRSVGRDLDDGDDLDVDAVGAAGAAILLSSRPVHLTPSLEVTPOAIS 660
Qy 661 VVQDCRRRQGEAVCLTAALCFQVTSRTPGRWHDQFYMRFTASLDEWTAGARAAFDGSGQ 720
Db 661 VVQDCRRRQGEAVCLTAALCFQVTSRTPGRWHDQFYMRFTASLDEWTAGARAAFDGSGQ 720
Qy 721 RLSPRRLRLSVGNVTCQLHFFHVLDTSDYLRPVALVTTFALDNTTKPGPVINEGSPTSIQ 780
Db 721 RLSPRRLRLSVGNVTCQLHFFHVLDTSDYLRPVALVTTFALDNTTKPGPVINEGSPTSIQ 780
Qy 781 KLVPFSDKCGPDNECVTDLVQNMIDIRGSRKAPFVVRGGRRKVLVSTTLNKENAYNT 840
Db 781 KLVPFSDKCGPDNECVTDLVQNMIDIRGSRKAPFVVRGGRRKVLVSTTLNKENAYNT 840
Qy 841 SLSLIFSRNLHLASLTQRESPIKVECAAPSAHARLCSVGHVPFQTKAKVTFLLEPFSC 900
Db 841 SLSLIFSRNLHLASLTQRESPIKVECAAPSAHARLCSVGHVPFQTKAKVTFLLEPFSC 900
Qy 901 SSSLSQVFGKLTASSDSLERNGTLOBNTAOTSAYIQVEPHLLFSSSE9TLHRYEVHPYGT 960
Db 901 SSSLSQVFGKLTASSDSLERNGTLOBNTAOTSAYIQVEPHLLFSSSE9TLHRYEVHPYGT 960
Qy 961 PVGPGPEFTKTLRVNLGCVYVSGLLISALLPAVAGHGNFYLSQVITNNASCIVQNLT 1020
Db 961 PVGPGPEFTKTLRVNLGCVYVSGLLISALLPAVAGHGNFYLSQVITNNASCIVQNLT 1020
Qy 1021 EPPGPPVHPHELOQTNRLNGSNTQCVVRCHLGQLAGKTEVSGLLRLVHNEPFRRAKFK 1080
Db 1021 EPPGPPVHPHELOQTNRLNGSNTQCVVRCHLGQLAGKTEVSGLLRLVHNEPFRRAKFK 1080
Qy 1081 SLTVVSTFELTEGSEVLQTEASRWSESLLEVVTQTPRILISLWILGSLGALLLALL 1140
Db 1081 SLTVVSTFELTEGSEVLQTEASRWSESLLEVVTQTPRILISLWILGSLGALLLALL 1140
Qy 1141 VFCLWKLGFPAHKKIPREEKKEEKLQ 1167
Db 1141 VFCLWKLGFPAHKKIPREEKKEEKLQ 1167
```

## RESULT 2

US-10-741-601-532

; Sequence 532, Application US/10741601

; Publication No. US20040166519A1

; GENERAL INFORMATION:

```
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; STENOSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 532
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-10-741-601-532
```

```
Query Match 99.8%; Score 6092; DB 16; Length 1177;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1164; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MELPFFVTHLPLVFLVTGLCSPFNLDHHPRLPPGPEAEFGYSVLQHVGGQRMWLVGA 60
Db 1 MELPFFVTHLPLVFLVTGLCSPFNLDHHPRLPPGPEAEFGYSVLQHVGGQRMWLVGA 60
Qy 61 PWDGPGSDRRGDVYRCFVGGAHNAPCAKGLHDYQLGNSHSPAVNMHLGMSLLETGDDG 120
Db 61 PWDGPGSDRRGDVYRCFVGGAHNAPCAKGLHDYQLGNSHSPAVNMHLGMSLLETGDDG 120
Qy 121 FMACAPLWSRACSSVFSSGICARVDASFPQGS LAPTAQRCPTYMDVVIVL DGSNSIYP 180
Db 121 FMACAPLWSRACSSVFSSGICARVDASFPQGS LAPTAQRCPTYMDVVIVL DGSNSIYP 180
Qy 181 WSEVQTLRLRLVGLFIDPEIQVGLVQGESPVHWSLGDFTKEEVVRAAKNLSRREG 240
Db 181 WSEVQTLRLRLVGLFIDPEIQVGLVQGESPVHWSLGDFTKEEVVRAAKNLSRREG 240
Qy 241 RETKTAQAIWVACTEGFSQSHGGPEARLLVVVTDGSHDGBELPAALKACAGRVTRY 300
Db 241 RETKTAQAIWVACTEGFSQSHGGPEARLLVVVTDGSHDGBELPAALKACAGRVTRY 300
Qy 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDRPFNFVNTDEAALTDIVDALGDRIFGLEGS 360
Db 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDRPFNFVNTDEAALTDIVDALGDRIFGLEGS 360
Qy 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGYDVGSGVLMLEGGHRLFPFRMALEDEF 420
Db 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGYDVGSGVLMLEGGHRLFPFRMALEDEF 420
Qy 421 PPALQNHAAVLYGYSVSMLLRGGRRLFLSGAPRFRHGKVIAPQLKKGAVRVAQSLQGE 480
Db 421 PPALQNHAAVLYGYSVSMLLRGGRRLFLSGAPRFRHGKVIAPQLKKGAVRVAQSLQGE 480
Qy 481 QIGSYFGSELCPDLTDGDTTDLVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLOQTLO 540
Db 481 QIGSYFGSELCPDLTDGDTTDLVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLOQTLO 540
Qy 541 PEPPQDARFGFAMGALPDNLQDGFADVAVGAPLEDHGQALYLYHGTSQSVRPHPAQRIA 600
Db 541 PEPPQDARFGFAMGALPDNLQDGFADVAVGAPLEDHGQALYLYHGTSQSVRPHPAQRIA 600
Qy 601 AASMPHALSYFGRSVGRDLDDGDDLDVDAVGAAGAAILLSSRPVHLTPSLEVTPOAIS 660
Db 601 AASMPHALSYFGRSVGRDLDDGDDLDVDAVGAAGAAILLSSRPVHLTPSLEVTPOAIS 660
Qy 661 VVQDCRRRQGEAVCLTAALCFQVTSRTPGRWHDQFYMRFTASLDEWTAGARAAFDGSGQ 720
Db 661 VVQDCRRRQGEAVCLTAALCFQVTSRTPGRWHDQFYMRFTASLDEWTAGARAAFDGSGQ 720
Qy 721 RLSPRRLRLSVGNVTCQLHFFHVLDTSDYLRPVALVTTFALDNTTKPGPVINEGSPTSIQ 780
Db 721 RLSPRRLRLSVGNVTCQLHFFHVLDTSDYLRPVALVTTFALDNTTKPGPVINEGSPTSIQ 780
Qy 781 KLVPFSDKCGPDNECVTDLVQNMIDIRGSRKAPFVVRGGRRKVLVSTTLNKENAYNT 840
Db 781 KLVPFSDKCGPDNECVTDLVQNMIDIRGSRKAPFVVRGGRRKVLVSTTLNKENAYNT 840
```

```
Qy 841 SLILFSRNHLASLTQRESPIKVECAAPSAHARLCSVGHVPFQTKAKVTFLEPFESC 900
Db 841 SLILFSRNHLASLTQRESPIKVECAAPSAHARLCSVGHVPFQTKAKVTFLEPFESC 900
Qy 901 SLLSQVFGKLTASSDLSLNGTLQENTAOISAYIQVEPHLLFSSESTLHRYEHPYGTLL 960
Db 901 SLLSQVFGKLTASSDLSLNGTLQENTAOISAYIQVEPHLLFSSESTLHRYEHPYGTLL 960
Qy 961 PVGPGPEFKTTLRVQNLGCVYVGLIISALLPAVAHGNFYLSQVITNASCIVQNL 1020
Db 961 PVGPGPEFKTTLRVQNLGCVYVGLIISALLPAVAHGNFYLSQVITNASCIVQNL 1020
Qy 1021 EPPGPPVHPELOHTNPLNGSNTOCQVVRCHLQOLAKGTESVGLLRLVHNEFPRAKFK 1080
Db 1021 EPPGPPVHPELOHTNPLNGSNTOCQVVRCHLQOLAKGTESVGLLRLVHNEFPRAKFK 1080
Qy 1081 SLTVVSTFELGTGEGSVLQTEASRWSESLLEVVTOTRPIILISLWILGSLGLLLALL 1140
Db 1081 SLTVVSTFELGTGEGSVLQTEASRWSESLLEVVTOTRPIILISLWILGSLGLLLALL 1140
Qy 1141 VFCLWKLGFPAHKKIPEEKREKLEQ 1167
Db 1141 VFCLWKLGFPAHKKIPEEKREKLEQ 1167
```

## RESULT 3

```
US-10-291-265-338
; Sequence 338, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 338
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-338
```

```
Query Match 40.08; Score 2439.5; DB 15; Length 1188;
Best Local Similarity 43.26; Pred. No. 1.1e-206;
Matches 517; Conservative 213; Mismatches 426; Indels 45; Gaps 16;
```

```
Qy 1 MELPFTVTHLPLVFLTGLCSPPNLDHHPRLPFPPEAEFGYSVLQVHGCGRWLVGA 60
Db 1 MDLPRGLVVAWALSFWGFTDTFMDTRKPRVIFGSRAPFGYTVQGHDSGNKWLVGA 60
Qy 61 PWGPGSDRRGDVYRCVPGGAHNAAPCAKHLGDLQYQNGNSHPAVNMHLGMSLLETGDGG 120
Db 61 PLETNGYQKTDVYKCPV---IHGNCCKNLNLRVTLNVSERKDNMRGLSLATNPKNDS 117
Qy 121 FMACAPLWSRACGSSVSSGICARVDASFQPGSLATPQRCPTMYDMWVILDCSNIYP 180
Db 118 FLACSLWSHCEGSSYTTGMCNSVNSFRSKTVAPALQRCQTYMDIVIVLDCSNIYP 177
Qy 181 WSEVOTFLRLRLVGLKFLTDPEQIQVGLVQYGSVPHEWSLGDFTKKEVWRAAKNLSREG 240
Db 178 WVEQVHFLINILKFKYIGPGQIQGVVQYGVQEDVVEHFLNDIRSKDVVEAASHIEBGG 237
```

## RESULT 4

```
US-10-291-265-810
; Sequence 810, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
```

```
Qy 241 RETTAAQIMVACTEGFSQSHGGRPEAARLLVVVTDGESHGDELPALAKACAGRVTRY 300
Db 238 TETRTAFIEFARSEAPQK--GGRKGAKKVMIVITDGHSDSPDLEKVIQOOSRDNTRY 295
Qy 301 GIAVLGHLRRQRDPSPSFLREIRTIASDPDEREFNFVNTDEAALTDIVDALGDRIFGLEGS 360
Db 296 AVAVLGYNRRGINPETFLNEIKYIASDDPDKHFFNVTTDEAALKDIDVALGDRIFSLEGT 355
Qy 361 HAENESSFGLMEQIQFSTHRLKDGILFGMWGAYDGGSVLWLEGHRLPPRMALEDEF 420
Db 356 N-KNETSFGLMEQIQFSSHVVEDGVLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEP 414
Qy 421 PPALQNHAAVLGYSVSSMLLRGGRRLFLSGAPRFRHGKVIAFOLKXGAVRVAQSLQGE 480
Db 415 PSELKHGALVGTIVTSSVSSROGRVAVAGAPRNFHTGKVLFTMHNRRSLTTHQAWRGQ 474
Qy 481 QIGSYFGESELCPLDTRDGTDLVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLOQTLLQ 540
Db 475 QIGSYFGESEITSDIDGDTVLLVGAPWYFN--EGRERKVVYVEL--RQNRVYVNGTLK 532
Qy 541 PEPP--QDARPGFAMGALPDINQDGFADVAGAPLEDHGHQALYLYHGTQSGVRPHPAORI 599
Db 533 DSHSYQNAREGSSITASVRDLNQDSYNDVVVVGAPLEDNHAGAIYIFHGFSGSILKTPQRI 592
Qy 600 AAASMPHALSYFGRSVDRLDDGDDLVDVAVGAQGAAILLSRPIVHLTPSLEVTPOAI 659
Db 593 TASELATGLQYFGCSIHGQDLNEDGLIDLAVGALGNVILWRPVPVQINASLHFEPSKI 652
Qy 660 SVVQDCRRRGOBAVCLTAALCFQVTSRTPGRDHGFQYMRFTASLDSEWTAGARAAPFGSG 719
Db 653 NIFHRDCKSRGRDATCLAAFLCTPTPLAPHQFTTVGIRYNATMDERRYTTPRAHDEGG 712
Qy 720 QRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRLPVALTVTTFALDNTTKPGVNLGSPSTSI 779
Db 713 DRETNRVALLSSGOELCERINFHVLDTADYVVKPVTFSVEYSLEDPDH--GPMLDDGWPTTL 771
Qy 780 QKLVPFSKDCGPDNECVTDLVLQVNMDI-----RGRK-----APVV 817
Db 772 RVSPVFMWNGCNEDEHCVDFDLVDARSDLPTAMEYQQRVLRKPAQDCSAYTLSFDTTTFII 831
Qy 818 RGGRRKVLVSTTLNKRKENAYNTSLISFNRNLHSLTLPQRESPIKVECAAPS--AHAR 875
Db 832 ESTQRVAVATEATLENGENAYSTVLNISQSANLQFASLIQKEDSDGSECVNEERRLQKQ 891
Qy 876 LCSVGHVPFQTKAKVTFLEPFESCSLLSOVFKLTASSDLSLERNGTQLOENTAOISAYI 935
Db 892 VCNVSPVFFRAKAKVAPRLDPFESKSIPLHLELELAAGSDSNERNSTKEDNVAPLRFHL 951
Qy 936 QYEPHLLFSSESTLHRYEHPYGTLP--VGPGEFETTLRVQNLGCVYVSGLLIISALLPA 993
Db 952 KYEADVLFTRSSLSLHYEVKPNSSLERYDGGPPFCIFRIQNLGLPFIHGMKMITIPI 1011
Qy 994 VAHGNYFLSLSQVITN--NASC--IVONLTPEPPPVHPELOHTNPLNGSNTOCQVVR 1050
Db 1012 ATSGNRLKLRPLDTEANTSCINWGNSTERYPTVE--EDLRAPQLNHSNDSVVSINC 1070
Qy 1051 HLGQAKGTESVGLLRLVHNEFFRRAKFKSLTVVSTFELGTGEGSVLQTEASRWSESL 1110
Db 1071 NI--RLVNQEIHFLLGNLWLSLKALKYKSMKIMVNAALQRPQSPFPIREEDPSQIV 1129
Qy 1111 LEVVQTRPILISLWILGSLVGLLLALLVFLWKLGLGFAHKKIPEEKREKLE 1166
Db 1130 FEISKQBDWQVPIIIVGSTLGGLLALLVALLWKLGLGFFRSAR----RRREPGLD 1181
```

```
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 810
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-291-265-810

Query Match
  39.9%; Score 2438.5; DB 15; Length 1188;
Best Local Similarity 43.2%; Pred. No. 1.4e-206;
Matches 517; Conservative 213; Mismatches 421; Indels 45; Gaps 16;

Qy 1 MELPFTVTHLFLPLVLTGLCSFNLDEHHPLRPPGPEAFGYSVLQHVGGQRMVLVGA 60
Db 1 MDLPRGLVAVAWLSLWPGFTDTFNMTRKERVIPGSRTPAFGYTVQOHDISGNKMLVGA 60
Qy 61 PWGPGSDRGDDVYRCVGGAHNAPCAKAGHLGQYDLGNSSHPAVNMHLGMSLLETDDGG 120
Db 61 PLETNGYQKTGDVYKCPV---IHGNCCKLNLGRVTLNSVSEKDNMRGLSLATNPKN 117
Qy 121 FMACAPLWSRACGSSVFSSGICARVDASFOQGSIAPTAQRCTPTMDVIVLQGSNIYP 180
Db 118 FLACSPLWSHCEGSSYTTGMSRVNSNFRFSKTVAPALQRCQTYMDIVIVLQGSNIYP 177
Qy 181 WSEVQTLRLRLVKGFLIDPQIQVLQYGESPVHWSLGDPRTEKEVVRVRAAKNLSREG 240
Db 178 WVEVQHELNLKFKYIGPGQIQGVQYGEDVVHPEFLNDYRSVDVVEAASHIEQGG 237
Qy 241 RETTAQAIWVACTEGSQSHGGRPEARLLVVTVDGESHGDELPAALKACAGRTY 300
Db 238 TETTAFGIEFARSEAFQK--GGRKAGAKVMIVITDGHSDSPLEKVIQOOSRDNTRY 295
Qy 301 GIAVLGHLRRORPSFLREIRTIASDPERFFNVNTDEAALTIDVDALGDRIFGLEGS 360
Db 296 AVAVLGYNNRGINPETFLEIKYIASDDPKHFFNTDEAALKDIDVDALGDRIFSLEGT 355
Qy 361 HAENESSFGLMSQIGSFTHRLKDGILLFGMVGYDMGGSVLMLEGHRLPPPRMALED 420
Db 356 N-KNETSFGLMSQIGSFTHRLKDGILLFGMVGYDMGGSVLMLEGHRLPPPRMALED 414
Qy 421 PPALONHAAYLGYSVSSMLLGRGRLFLSGAPRFRHKGKVIAPQLKDGAVRVAQSLQGE 480
Db 415 PEEUNHGAUYLGTVTVSVSSRQGRVYVAGAPRPNHTGKVIPLTFMHNRSLLTHQAMRGQ 474
Qy 481 QIGSYFSGELCPDTRDGTDLVLLVAAPFLGPQNKETGRVYVVLVGGQSLTLQGTLO 540
Db 475 QIGSYFSGELTSDVIDGGVTDVLLVGAAPYFN--EGRGRGVYVEL--QNRFPYNGTLK 532
Qy 541 PEPP-QDARFGAMGALPDNLQDGFADVAGAPLEDGHQALYLYHGTQSGVRPHPAQRI 599
Db 533 DSHSYQNAFGSSIASVRDLNQDSYNDVWVAGAPLEDNHAGAIYIFHGFSGSILKTPQRI 592
Qy 600 AAASMPHALSYFSGSVDRDLDDGLVDVAVAGQAAILLSRPVHLFSPLEVTQAI 659
Db 593 TASELATGLQYFGSIHQDLNEDGLIDLAVGALNAVLMSRPVQINASLHFEPSKI 652
Qy 660 SVVORDCRRRQGEAVCLTAALCFQVTSKTPGRWDHQFVMRFTASLDSEWTAGARAAPDGG 719
Db 653 NIFHRDCKSRDRTCLAAFLCTFPIFLAHPFQTTTIGIRYNATMDERRYTTPRAHLEGG 712
Qy 720 QRLSPRLRLSVGNVTCQLHFHVLDTSDYLRPVALTVTFALDNTTKPGPVNLGSGSTSI 779
Db 719 QRLSPRLRLSVGNVTCQLHFHVLDTSDYLRPVALTVTFALDNTTKPGPVNLGSGSTSI 779
```

```
Db 713 DRFTNRAVLLSSGOELCERINFHVLDTADYVKPVTFSEVSESLDDPDH--GPMLLDDGWPTTL 771
Qy 780 QKLVFPFSKCGDNECVTDVLQVNMDI-----RGSRK-----APFVV 817
Db 772 RVSVFPWNGCNEDEHCVPLVDLARSDLPTAMEYCORVLURKPAQDCSAVTLSPDTTVFII 831
Qy 818 RGRRKVLVSTTLNENKENAYNTSLIIFSRNLHLASLTPQRESPIKVECAAPS--AHAR 875
Db 832 ESTRQRAVEATLENKENAYSTVLNISQSANLQFASLIQKEDSDGSIQVNEERLQKQ 891
Qy 876 LCSVGHVPQTGAQVTFLLPEFSCSSLSQVFGKLTASDSLSLERNGTQOENTAQTSAYI 935
Db 892 VCNVSYFPFRKAKAVAFRLDFEFKSIIFLHLEIAGSDSNRSTKEDNVAPLRFHL 951
Qy 936 QYEPHLLFSSESTLHRYEVHPTLP--VGPGEPTTLRVQNLGCVVSGLSLALLPA 993
Db 952 KYEVDVLTFRSSLSHYEVKPNSSRLERYDGIQPPFCIFRIQNLGLFPIHGMKMITIPI 1011
Qy 994 VAHGGNYFLSQTITN--NASC-IVQNLTEPPGPPVHPPEELQHTNRLNGSNITCQCVRC 1050
Db 1012 ATRESGNRLKLRDLTDEANTSCNINWSTERYPTVE--EDLRRAPQLNHSNSDVVSINC 1070
Qy 1051 HLGQAKAGTEVSGLLRLVHNEFFRRAKFKSLTVVSTFELGTEEGSVLQITEASRWSL 1110
Db 1071 NI-RLVPNOEINPHLLGNLWLSLALKYKSMKIMVNAALQROFHSPFFIFREEDPSRQIV 1129
Qy 1111 LEVVQTRPILISLWILIGSVLGLLALLLVCLWKLGFPAHKKIPEEKKREKLE 1166
Db 1130 FEISKQEDMQVPIIIVGSTLGLGLLALLLVCLWKLGFPAHKKIPEEKKREKLE 1181

RESULT 5
US-09-984-130-35
; Sequence 35, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-984-130-35

Query Match
  39.8%; Score 2429; DB 10; Length 1189;
Best Local Similarity 43.1%; Pred. No. 9.8e-206;
Matches 516; Conservative 213; Mismatches 422; Indels 46; Gaps 16;

Qy 1 MELPFTVTHLFLPLVLTGLCSFNLDEHHPLRPPGPEAFGYSVLQHVGGQRMVLVGA 60
Db 1 MDLPRGLVAVAWLSLWPGFTDTFNMTRKERVIPGSRTPAFGYTVQOHDISGNKMLVGA 60
Qy 61 PWGPGSDRGDDVYRCVGGAHNAPCAKAGHLGQYDLGNSSHPAVNMHLGMSLLETDDGG 120
Db 61 PLETNGYQKTGDVYKCPV---IHGNCCKLNLGRVTLNSVSEKDNMRGLSLATNPKN 117
Qy 121 FMACAPLWSRACGSSVFSSGICARVDASFOQGSIAPTAQRCTPTMDVIVLQGSNIYP 180
Db 118 FLACSPLWSHCEGSSYTTGMSRVNSNFRFSKTVAPALQRCQTYMDIVIVLQGSNIYP 177
```



Db 118 FLACPLWSHECCSSYYTTCMSRVNSFRFSKTVAPALQRCQTYMDIVVLGDSNIYP 177  
Qy 181 WSEVQTFRLRLVKGKLFIDPEQIQVGLVQYGESPVHWSLGDFTKBEVVAANKLSRREG 240  
Db 178 WVEVQHPLINILKFKVIGPQIQGVVQYGEDVHVEPHLNDYRSVDVVEAASHIEQGG 237  
Qy 241 RETKTAQAINVACTEGFSQSHGGRPEAARLLVVVTGESHGDELPAALKACAGRVTRY 300  
Db 238 TETRTAFGEFARSEAFQK--GGRKGAKYMIIVITGESHSDSPLEKVIQOOSRDNVTRY 295  
Qy 301 GIAVLGHLRRQDPSPFLREIRTIASDPDERFFNVNTDEAALTDIVDALGDRIFGLEGS 360  
Db 296 AVAVLYGNRRGINPETFLNEIKYIASDDPKHFFNVTDAAALKDVIDALGDRIFSLEGT 355  
Qy 361 HAENESSFGLEMSQIGFSTHRLKDGILFGMVGYDNGSVLWLEGGHRLFPFRMALEDEF 420  
Db 356 N-KNETSFGLMSQTFSSHVEDVLLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEF 414  
Qy 421 PPALQNHAAVLYGSVSMLLRGRRLLFLSCGAPFRHKGKVIATFOLKKGAVRVAQSLQGE 480  
Db 415 PEEKKHGAYLGYTVTSVSWSSROGRVYVAGAPRPHHTGKVLFTMHNRRSLTIHQAMRGQ 474  
Qy 481 QIGSYFGSELCLPTDRDGTDTLLVAAPMFLGPQNKETGRVYVYLWQOQSLTLTQGLTQ 540  
Db 475 QIGSYFGSEITSDVDIGDGTVDLLVGAPMYFN--EGRERGKVVYVEL--RQNRFYVNGTLK 532  
Qy 541 PEPP-QDARFGFAMGALPDNLQDGFADVAVGAPLEDGHQAGLYLHYGTQSGVRPHPAQRI 599  
Db 533 DSHSYQNAFGSSIASVRDLNQDSYNDVVVVGAPLEDNHAGAIYIFHGFGRSILKTPKQRI 592  
Qy 600 AAASMPHALSYFGRSVDGRLDGLDGLDVLVDVAVGAQGAAILLSRPPIVHLTPSLEVTQAI 659  
Db 593 TASELATGLQYFGCSIHGOLDNEDGLIDLAVGALGNVILWSPVQVQINASHLFEPSKI 652  
Qy 660 SVVQDCCRERGOBAVCLTAALCFQVTSRTGRWDHQFYMRFTASLDSEWTAGARAAFDGSG 719  
Db 653 NIHRDCKRSGRDATCLAAFLCTPIPLAFHFQTTTGVIRYNATMDERRITPRAHLDGEG 712  
Qy 720 QRLSPRLRLSVGNVTCQELHFHVLDTSDYLRPVALTFTFALDNTTKPGVJNAGSPSTI 779  
Db 713 DRFTNRAVLSSGOELCERINFHVLDTADYVVKPVTFSVEYSLEDPDH--GPMLDGDMWPTTL 771  
Qy 780 QKLVPSKDCPDNECVTDVLQVNMDDI-----RGRK-----APFVV 817  
Db 772 RVSPFWMGNCNEDEHCVDPDLVLDARSOLDPTAMEYCCQRLVRKPAQDCSAYTSLSDTTFVFI 831  
Qy 818 RGRKRLVSTTLNKRKENAYNTSLIIFSRNLHLASLTQRESPIKVECAAPS--AHAR 875  
Db 832 ESTRQORVAVEATLENGENAYSTVNLISQANLOFASLIQKEDSDGSEICWNERRLQKQ 891  
Qy 876 LCSVGHVPVQTKAVTFLLEBFSCSLLSOVFGKLTASSDSLERNGTLOENTAQTSAYI 935  
Db 892 VCNVSYFPFRAKAVAFRLDPEFSKSIFLHLETELAAAGSDSNERDSTKEDNVAPLRFHL 951  
Qy 936 QYEPHLLFSSSESTLHREYVHPYGTLP--VGPGEFKTTLRVONLGCYVSVGLIISALLPA 993  
Db 952 KYEADVLTFRSSLSHYEVKLSLERYDYGPPFCIFRIQNLGLFPFIHGMKMITIPI 1011  
Qy 994 VAHGNYFLSLQSVITN--NASC-IVONLTERPPGVPHPELQHTNRLNGSNTOCCVVR 1049  
Db 1012 ATRSNRLKLRLDFLTDDEVANTSCNIWGNSTEYRPTVE--EDLRAPQLNHSNDVVSIIN 1070  
Qy 1050 CHLGQLAGKTEVSVGLLRLVHNEFFRRAKFSITVVTSTFELGTBEGSVLQLEASRWSES 1109  
Db 1071 CNI-RLVPNQEIHFLLGNLWRLSLKALKYKSMKIMVNAALQRFHSPFIFREEDPSRQI 1129  
Qy 1110 LLEVQTRPILISWILIGSVGLLILALLVFCWLKGLGFPAHKKIPIEEKREBLE 1166  
Db 1130 VFEISKQEDQVPTIIVTGSTGLGLLLALLVIALWKLGFPRFAR----RRREPLD 1182

RESULT 6  
US-09-836-353A-35

; Sequence 35, Application US/09836353A  
; Publication No. US20030129685A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 12 Human Secreted Proteins  
; FILE REFERENCE: PP489P1  
; CURRENT APPLICATION NUMBER: US/09/836,353A  
; CURRENT FILING DATE: 2001-04-18  
; PRIOR APPLICATION NUMBER: 60/198,407  
; PRIOR FILING DATE: 2000-04-19  
; PRIOR APPLICATION NUMBER: PCT/US99/25031  
; PRIOR FILING DATE: 1999-10-27  
; PRIOR APPLICATION NUMBER: 60/105,971  
; PRIOR FILING DATE: 1998-10-28  
; NUMBER OF SEQ ID NOS: 147  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 35  
; LENGTH: 1189  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-836-353A-35  
  
Query Match 39.8%; Score 2429; DB 10; Length 1189;  
Best Local Similarity 43.1%; Pred. No. 9.8e-206;  
Matches 516; Conservative 213; Mismatches 422; Indels 46; Gaps 16;  
  
Qy 1 MELPFTVTHLPFLPVLTGLCSFPNLDHHPRLPPGPEAEFGYSVLQHVGGQRMVLVGA 60  
Db 1 MDLPRLGLVAVWALSLLWPGFTDTFNMOTRKPRVTPGSRTAFFGYTVQQHDISGNKWLTVGA 60  
Qy 61 PWDGPGSDRRGDVYRCPVGGAHNAPCAKHGLGDYQLGNSHPAVNHLGMSLLETDDGG 120  
Db 61 PLETNGYQKTGDVYKPCV---IHGNCCTKLNLGRVTTLSNVSERKDNMRLGLSLATNPKNDS 117  
Qy 121 FMACAPLWSRACSSVFSFGICARVDASFOQGLSAPTAQRCPTYMDVWLVDGNSNIYP 180  
Db 118 FLACPLWSHECCSSYYTTCMSRVNSFRFSKTVAPALQRCQTYMDIVVLGDSNIYP 177  
Qy 181 WSEVQTFRLRLVKGKLFIDPEQIQVGLVQYGESPVHWSLGDFTKBEVVAANKLSRREG 240  
Db 178 WVEVQHPLINILKFKVIGPQIQGVVQYGEDVHVEPHLNDYRSVDVVEAASHIEQGG 237  
Qy 241 RETKTAQAINVACTEGFSQSHGGRPEAARLLVVVTGESHGDELPAALKACAGRVTRY 300  
Db 238 TETRTAFGEFARSEAFQK--GGRKGAKYMIIVITGESHSDSPLEKVIQOOSRDNVTRY 295  
Qy 301 GIAVLGHLRRQDPSPFLREIRTIASDPDERFFNVNTDEAALTDIVDALGDRIFGLEGS 360  
Db 296 AVAVLYGNRRGINPETFLNEIKYIASDDPKHFFNVTDAAALKDVIDALGDRIFSLEGT 355  
Qy 361 HAENESSFGLEMSQIGFSTHRLKDGILFGMVGYDNGSVLWLEGGHRLFPFRMALEDEF 420  
Db 356 N-KNETSFGLMSQTFSSHVEDVLLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEF 414  
Qy 421 PPALQNHAAVLYGSVSMLLRGRRLLFLSCGAPFRHKGKVIATFOLKKGAVRVAQSLQGE 480  
Db 415 PEEKKHGAYLGYTVTSVSWSSROGRVYVAGAPRPHHTGKVLFTMHNRRSLTIHQAMRGQ 474  
Qy 481 QIGSYFGSELCLPTDRDGTDTLLVAAPMFLGPQNKETGRVYVYLWQOQSLTLTQGLTQ 540  
Db 475 QIGSYFGSEITSDVDIGDGTVDLLVGAPMYFN--EGRERGKVVYVEL--RQNRFYVNGTLK 532  
Qy 541 PEPP-QDARFGFAMGALPDNLQDGFADVAVGAPLEDGHQAGLYLHYGTQSGVRPHPAQRI 599  
Db 533 DSHSYQNAFGSSIASVRDLNQDSYNDVVVVGAPLEDNHAGAIYIFHGFGRSILKTPKQRI 592  
Qy 600 AAASMPHALSYFGRSVDGRLDGLDGLDVLVDVAVGAQGAAILLSRPPIVHLTPSLEVTQAI 659  
Db 593 TASELATGLQYFGCSIHGOLDNEDGLIDLAVGALGNVILWSPVQVQINASHLFEPSKI 652  
Qy 660 SVVQDCCRERGOBAVCLTAALCFQVTSRTGRWDHQFYMRFTASLDSEWTAGARAAFDGSG 719  
Db 653 NIHRDCKRSGRDATCLAAFLCTPIPLAFHFQTTTGVIRYNATMDERRITPRAHLDGEG 712

```
Qy 720 QRLSPRRLRLSVGNVTCQELHFVHLDTSYLRPVALVTFTALDNTTKPGPVLMNEGSPTSI 779
Db 713 DRFTNRAVLSSGQELCERINFHVLDVADYVVKVPTSEVSLBDDPDH-GPMLDGDGWPTTL 771
Qy 780 QKLVPPSKDCGPDNECVTDVLQVNMDI-----RGRK-----APFVV 817
Db 772 RVSPVPPNGCNEDEHCVPLDVLDAARSLPTAMEYCQRLRKPAPQDCSAYTSLSFDTTVFII 831
Qy 818 RGRKRVLVSTTLNRKENAYNTSLIIPGRNLHLASLTPORESPIKVECAAPS--AHAR 875
Db 832 ESTQRVAVATLENRENAYSTVLNISQSANQFASLIQKEDSDGSEICVNEERLQKQ 891
Qy 876 LCSVGHVPFGAKVTFLLEBFSCSSLLSQVKLTASSDSLSERNGTLOENTAQTSAIY 935
Db 892 VCNVSYPPFRAKAVAFRLOFEPFSKIFLHLEIELAAGSDSNERDSTKEDNVAFLRFL 951
Qy 936 QYEPHLLFSSESTLHYREVPYGTLP--VQPGPEFKTTLRVQNLGCVVSGLIISALLPA 993
Db 952 KYEADVLTFRSSLSHYEVKLSLERYDYGPPFCIFRIONLGLFPFIHGMKTIPI 1011
Qy 994 VAHGNNYFLSLSQVITN--NASC-IVONLTETPPGPPVHPEELQHTNRLANGSNTQCQVR 1049
Db 1012 ATRSGNLLKLDRFLTDVANTSCNMGNSTEYRPTVE-EDLRRAPQLNHSNDVVVSIN 1070
Qy 1050 CHLQGLAKGTEVSVGLLRLVHNEFFRAKFKSLTVSTFELGTBEGSVLQLTASRWSES 1109
Db 1071 CNI-RLVPNGEINPHLLGNLWRLSLKALKYKSMKIMVNAALQRFHSPFIFREDDPSROI 1129
Qy 1110 LLEVQTRPLIISWILIGSVLGGLLALLLVFCMLKGLGFAHKKIPBEKREBLE 1166
Db 1130 VFEISKQEDMQVPTIITVGTSLGGLLLALLLVLMKLGFRSAR----RRRSPGLD 1182

RESULT 7
US-10-262-839-4
; Sequence 4, Application US/10262839
; Publication NO. US20040038877A1
; GENERAL INFORMATION:
; APPLICANT: Alebrook, John,
; APPLICANT: Anderson, David W.,
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Burgess, Catherine,
; APPLICANT: Catterton, Elina,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Ji, Weizhen,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Smithson, Glennda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Taupier, Raymond, Jr.,
; APPLICANT: Vernet, Corine,
; APPLICANT: Voss, Edward,
; APPLICANT: Zerhusen, Brian,
; APPLICANT: Zhong, Mei,
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-462A
; CURRENT APPLICATION NUMBER: US/10/262,839
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
```

```
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,101
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/327,342
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 60/328,044
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/374,738
; PRIOR FILING DATE: 2002-04-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 4
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-262-839-4

Query Match 39.8%; Score 2429; DB 15; Length 1189;
Best Local Similarity 43.1%; Pred. No. 9.8e-206;
Matches 516; Conservative 213; Mismatches 422; Indels 46; Gaps 16;

Qy 1 MELPFFVTHLPLVLFTGLCSPPNLDHHPRLPPGPPREAFYGVSLQHVGGQGRWMLVGA 60
Db 1 MDLPRLGVAVAWALLSWPGFTDTFNTMDTRKPRVPVPGSRTAFPGFTVQOQHDISGNKWL VGA 60
Qy 61 PWDGSDRGDGVYRCVCGAHNAPCAKGLHDYQLGNSHPAVNMHLGMSLLETDDGG 120
Db 61 PLESTNGYQKTGDVYKCPV---IHGNCCTKLNGLVTLNSVSEKDNRLGLSLATNPKN 117
Qy 121 FMACAPLWSRACSSVFSFGICARVDASFPQGSGLAPTACRQCTPTYMDVVLVLDGNSIYP 180
Db 118 FLACSPLWSHECGSSVYTTGMCSSRVNSNFRSKTVAPALQRCQTYMDIVLDGNSIYP 177
Qy 181 WSEVQFRLRLVOKLFIDPEIQVGLVQVGSSEPVHWSLGDFTKEEVVPAANKLRSREG 240
Db 178 WVEVOHFLINILKKFYIGPQIQVGVVQGVGDEVHFLNDYRSVDVVEAAASHIQRGG 237
Qy 241 RETKTAQIMVACTEGFSQSHGGRPEARLLVVVTGESHGDBELPAALKACAGRVTRY 300
Db 238 TETRTAFGIEFARSEAFQK--GGRKGAKVMVITDGESHSDSPDLEKVIQOESRDNVTRY 295
Qy 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFFNVTDEAALTDIVDALGDRIFGLEGS 360
Db 296 AVAVLGYNNRGINPETFLNEIKYIASDPDDKHFFNVTDAAALDKDIVDALGDRIFSLEGT 355
Qy 361 HAENESSFGLMSQIGFSTHRLKXGILFGMVGYDMGSGSVLMLEGHRLPPPPMALEDEF 420
Db 356 N-KNETSFGLMSQIGFSSHVVVDGVLGAVGAYDNWGAVLKETSAGKVIPLRESYLKEP 414
Qy 421 PPALQNHAAVLYGYSVSMLLRGGRLLFLSGAPFRHRGVIAFOLKXDGAVRVAQSLQGE 480
Db 415 PEELKNHGAYLGYTVTSVSVSSROGRVYVAGAPFNHTGKRVLLFTMNNRSLTTHQAMRGG 474
Qy 481 QIGSYFGSELCPDLDTRDGTVDVLLVAAPMFLGPONKETGRVYVYLVVQGSLLTLQGTTLQ 540
Db 475 QIGSYFGSEITSVDIDGDTVDVLLVAGAPYFN-EGRERKGVVYVEL-RQNRFFVYNGTLK 532
Qy 541 PEPP-QDARFGFAMGALPDNLNQGFADVAVGAPLEDHGHQALYLYHGTSQGVPHPAQRI 599
Db 533 DSHSYQNARFSGSIASVRDLNQSDVNDVVVVGAPLEDNHAGAIYIFHGFRGSILKTPKQRI 592
Qy 600 AAASMPHALSYFGRSVDGLDDGDDLDVAVGACQAAIILSSRPVHLTPSLVETVPAI 659
Db 593 TASELATGLQYFGCSIHGQJDLNEDGLDILAVGALNAVILNSRPVVQINASLHFEPSKI 652
```

QY 660 SVVQDCRRRGOEAVCLTAALCFQVTSRTSGRWDHQFYMRTASLDSEWTAGARAAFGSG 719  
D 653 NIFHRDCKRGRDATCLAAFLCTPIFLAHFQVTTGIRYNATMDERRYPTRAHLDEGG 712  
QY 720 QRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALTFTFALDNTTKPGVNLNRSPTS 779  
D 713 DRFTNRAVLLSSGQELCERINFHVLDADYVVKPTFSVEYSLEDPDH-GPMLDDGWPTTL 771  
QY 780 QKLVFESKDCPDNECVTDVLQVNMID-----RGRK-----APFVV 817  
D 772 RVSVFPMWNGCNEDEHCVDPDLVLDARSDLPTAMEYCORVLRKPAQDCSAYTSLFDTTVPFI 831  
QY 818 RGRKRVLVSTTLNKRKENAYNTSLIIFSRNLHLASITPQRESPIKVECAAPS--AHAR 875  
D 832 ESTQRVAVATLENGENAYSTVLNISQSANLOFASLIQKEDSDGSECVNEERLQKQ 891  
QY 876 LCSVGHVPFOTGAKVTFLEFEFSCSLLSQVFKLTASDSLSLRNLTQENTTAQTSAYI 935  
D 892 VCNVSYPFRAKAKVAFRLDFEFKSIFLHLELELAAGSDSNERDSTKEDNVAPLRFHL 951  
QY 936 QYEPHLLFSSESTLHRYEHPYGTLP--VGPGEFETTLRVQNLGCVVSGLLIISALLPA 993  
D 952 KYEADVLFTRSSLSHYEVKLNLSLRYDGIKPPFSCIFRIONLGLPFIHGMKTIPI 1011  
QY 994 VAHGNVFLSLSQVITN--NASC-IVQNLTEPPGPPVHPEELOHTNRLNGSNTQCQVVR 1049  
D 1012 ATRSGNRLKLRLDRLTDEVANTSCNIWGNSTYPTPV--EDLRAPQLNHSNDVVSIN 1070  
QY 1050 CHLQOLAKGTEVSVGLRLVHNEFFRRAKFKSLTVSTFELGTREGSVLQITEASRWSES 1109  
D 1071 CNI-RLVPNGEINFHLLGNLWRLSLKALKYKSMIMVNAALQRFHSPFIREDPDSRQI 1129  
QY 1110 LLEVQTRPILISUWILIGSVLGGILLIALLVFCWLKGLGPPFAHKKIPEEKREKLE 1166  
D 1130 VFEISKQEDQVPIIIVGSTGLGILLIALLVLAALWKLGFPSAR----RRREPGLD 1182

RESULT 8  
US-09-984-130-43  
; Sequence 43, Application US/09984130  
; Publication No. US20030055231A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 12 Human Secreted Proteins  
; FILE REFERENCE: PF489P2  
; CURRENT APPLICATION NUMBER: US/09/984,130  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 60/243,792  
; PRIOR FILING DATE: 2000-10-30  
; PRIOR APPLICATION NUMBER: 09/836,353  
; PRIOR FILING DATE: 2001-04-18  
; PRIOR APPLICATION NUMBER: 60/198,407  
; PRIOR FILING DATE: 2000-04-19  
; PRIOR APPLICATION NUMBER: PCT/US99/25031  
; PRIOR FILING DATE: 1999-10-27  
; PRIOR APPLICATION NUMBER: 60/105,971  
; PRIOR FILING DATE: 1998-10-28  
; NUMBER OF SEQ ID NOS: 149  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 43  
; LENGTH: 1034  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-984-130-43

Query Match 36 9%; Score 2254; DB 10; Length 1034;  
Best Local Similarity 44.5%; Pred. No. 2.8e-180;  
Matches 464; Conservative 187; Mismatches 356; Indels 36; Gaps 11;

QY 1 MELPFVTHLFLPLVFLVGLCSPPNLDHHPRLFPFGPEAEFGYSVLQVHGCGRWMLVGA 60  
D 1 MDLPRLVVAWALSLWFGFTDTFNMDTKPRVIFGSRTAFFGYTVQDHDSGNKWLVGA 60

RESULT 9  
US-09-836-353A-43  
; Sequence 43, Application US/09836353A  
; Publication No. US20030129685A1

QY 61 PWDGSGDRRDYVRCVPVGAHNAPCAKGLHDYQLGNSSHPAVNMHGLMSLLETDDGPG 120  
D 61 PLETNGVQKTDGVTKCPV---IHGNC TKLNLGRVTLTNSVSRKDNMRGLSLATNPKN 117  
QY 121 FMACAPLWSRACSSSFSSGICARVDASFOFGSLAPTACRCPTYMDVIVILVDSGNSIYP 180  
D 118 FLACSPLWSHECGSSYYTTGMCSSRVNSFRFSKTVAPALQRCQTYMDIVILVDSGNSIYP 177  
QY 181 WSEVQTPRLRLVGLKFLDPEQIQVGLVQYGESPVHESLGDFTFYKEEVRAAKNLSRREG 240  
D 178 WVEYQHFELINILKXYIGPGQIQGVVQYGEDVVEHFLNDYRSVKDVVEAASHIEQRGG 237  
QY 241 RETKTAQINIVACTEGFSQSHGSRPEARLLVVVTDGESHGDELPALAKACAGRVTRY 300  
D 238 TETRTAFGIFAKSEAFQK--GGRKGAKVMIVITDGHSDSPLEKVIQOESRDNVTRY 295  
QY 301 GIAVLGHLRRQRPDPSSFLREIRTIASDPDERPFNVTDEAALTDIVDALGDRIFGLEGS 360  
D 296 AVAVLGYNNRRGINPETFLNEIKYIASDDDKHFFNVTDEAALKDIVDALGDRIFLSLEGT 355  
QY 361 HAENESSFGLMSQIGPSTHRLKDGILPGMWGAYDWGSGVLWLEGGHRLFPFRMALEDEF 420  
D 356 N-KNETSFGLMSQTGFSHVVEDGVLGAVGAYDMNGAVLAKETSAGKVIPLRESYLKEF 414  
QY 421 PPALQNHAAVLYGSVSMLLRGGRRLLFLSGAPRFRHKGKVIATFOLKKGAVRVVAQSLOGE 480  
D 415 PEELKNGHALGYTVTSVSSRQGRVYVAGAPRNFHTGKVLFTMHNRSSTIITQAMRGQ 474  
QY 481 QIGSYFSGELCLDTRDGTDTLLVAAPMFLGPQNKETGRVYVYLVQQSLLTQGLTQ 540  
D 475 QIGSYFSGEITSVDIDGQVTDVLLVGPAMYFN--EGRERKGVYVEL--RQNFVYNGTLK 532  
QY 541 PEP--QARFGFAMGALPDNLQDGFADVAVGAPLEDHOGALYLYHGTQSGVRPHPAQRI 599  
D 533 DSHSYQNAFSGSSIASVRDLNQDSYNDVAVGAPLEDNHAGAIYIFHGFSGSILKTPQRI 592  
QY 600 AAASMPHALSYFGRSVDRLDGDGLADVAVAGGAAILLSRPPIVHLTPSLEVTPOAI 659  
D 593 TASELATGLQYFGCSIHGQDLNEDGLIDLAVGALGNVILWSRPVVOINASLHFEPSKI 652  
QY 660 SVVQDCRRRGOEAVCLTAALCFQVTSRTSGRWDHQFYMRTASLDSEWTAGARAAFGSG 719  
D 653 NIFHRDCKRGRDATCLAAFLCTPIFLAHFQVTTGIRYNATMDERKRYTPTRAHLDEGG 712  
QY 720 QRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALTFTFALDNTTKPGVNLNRSPTS 779  
D 713 DRFTNRAVLLSSGQELCERINFHVLDADYVVKPTFSVEYSLEDPDH-GPMLDDGWPTTL 771  
QY 780 QKLVFESKDCPDNECVTDVLQVNMID-----RGRK-----APFVV 817  
D 772 RVSVFPMWNGCNEDEHCVDPDLVLDARSDLPTAMEYCORVLRKPAQDCSAYTSLFDTTVPFI 831  
QY 818 RGRKRVLVSTTLNKRKENAYNTSLIIFSRNLHLASITPQRESPIKVECAAPS--AHAR 875  
D 832 ESTQRVAVATLENGENAYSTVLNISQSANLOFASLIQKEDSDGSECVNEERLQKQ 891  
QY 876 LCSVGHVPFOTGAKVTFLEFEFSCSLLSQVFKLTASDSLSLRNLTQENTTAQTSAYI 935  
D 892 VCNVSYPFRAKAKVAFRLDFEFKSIFLHLELELAAGSDSNERDSTKEDNVAPLRFHL 951  
QY 936 QYEPHLLFSSESTLHRYEHPYGTLP--VGPGEFETTLRVQNLGCVVSGLLIISALLPA 993  
D 952 KYEADVLFTRSSLSHYEVKLNLSLRYDGIKPPFSCIFRIONLGLPFIHGMKTIPI 1011  
QY 994 VAHGNVFLSLSQVITNASCIV 1016  
D 1012 ATRSGNRLKLRLDRLTDEGEHVL 1034

```

GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
; OS-09-836-353A-43

Query Match          36.9%; Score 2254; DB 10; Length 1034;
Best Local Similarity 44.5%; Pred. No. 2,8e-190;
Matches 464; Conservative 187; Mismatches 356; Indels 36; Gaps 11;

QY 1 MELPFVTHLPLPLVLTGCLSPFNLDHHPLRFPQPPBAEFGYSVLQVHGQGRMLVGA 60
DB 1 MDLPRLVAVALLSSLPWPGTDTFNMDTRKPRVIPGSRTPAFGYTVQOHDISGNKWL VGA 60
QY 61 PWDGSGDRRDVYRCPUGANAPCAKGHLDYQLGNSSHPVAVNMHLCMSLLETDDGG 120
DB 61 PLETNQYQTDGYDKPV---IHGNCITKLNLGRVTLNSVSRKONMRLGLSLATNPKDNS 117
QY 121 FNACAPLWRACGSSVFSSGICARVDASPOQGS LAPTAQRCPTVMVVILVLDGNSIYP 180
DB 118 FLACSLPWHGCGSSYTTGMSRVNSNRPFKTVAPALQRCQTVMDIVIVLDGNSIYP 177
QY 181 WSEVQTFLLRLVGLKFLIDPEIQVLQVYGESPVHWSLGDPRTKBEYVRAAKNLSRREG 240
DB 178 WVEVQHFLINILKFKFYIGQIQGVQVQGEDVVHEFLNDVRSVKDVVEAASHIEQRGG 237
QY 241 RETTQAALVWACTEFGSOSHGRPEAARLLVVTDGSHDGEELPAALKACEAGRVTRY 300
DB 238 TETRTAFGIEFARSEAFQK--GGRKGAKKVMIVITDGHSDSPDLKVKVIOQSERDNVTRY 295
QY 301 GIAVLGHVLRQRDPSSFLREIRTTASDDPDERFFNVTDDEALTDIVDALGDRIFGLEGS 360
DB 296 AVAVLGYINRGINPFTFELNFKYIASDDPDDKHFNFVTDDEALDKDIVDALGDRIFSLGRT 355
QY 361 HAENESFGLMSQISGFSTRHLKDGLIFGMVAYDWSGVLMLEGGHRLFPFRMALEDEF 420
DB 356 N-KNETSFGLMSQISGFSSHVEDGVLLGAVGAYDMNGAVLKETAGKVIPLRSYLKEF 414
QY 421 PPALQNHAAVLYGSVSSMLLRGRRLFLSGAPFRHRKVIAPQLKKGAVRVVAQSLQGE 480
DB 415 PEBLKHGAYLYGTVTSVSSRQGRVYVAGAPFNHTGKVLFTMHNNRSLTIHQAMRGQ 474
QY 481 QIGSYFGLSCLPLDTRDCTTDLVLAAPMFLGPQNKETGRVYVYLVVQOSSLITQGLTQ 540
DB 475 QIGSYFGLSRLTSDIDGDTVDLLVAGAPMYFN--EGRSRKGVYVYEL--RQNRVYVNGFLK 532
QY 541 PEPP-QDARFGFAMGALPDNLQDGFADVAVGAPLEDGHGALYLYHGTSQGVRRHPAQR 599
DB 533 DSHSYQNARFGSIAVRDLNODSYNDVVVGAPLEDNHAGAIYIHPFGSILTKTPKRI 592
QY 600 AAASMPHALSYFGRSVDGRDLDDGLDVAVGAQAAILSSRPVHLTSPLEVTPOAI 659
DB 593 TASELATGLQYFGCSIHGQLDLNEDGLIDLAVGALGNVILMSRPVVOINASLHPEPSKI 652
QY 660 SVVQRCRRGQGAVALCTAALCFQVTSRTPGRWDHQFYMRFTASLDENTAGARAADGSG 719
DB 653 NIFHRDCKRSGRDATCLAAFLCPTPIFLAPHFPQTTTVGIRVYNATWDEKRYTPRAHLDEGG 712
QY 720 QRLSPRLRLSLVGNVTCEOLHFLHVLDTSDYLFVALTVTFALDNTTKPGPVLNKGSPS 779

```

; PRIOR APPLICATION NUMBER: 60/328,044  
 ; PRIOR FILING DATE: 2001-10-09  
 ; PRIOR APPLICATION NUMBER: 60/328,849  
 ; PRIOR FILING DATE: 2001-10-12  
 ; PRIOR APPLICATION NUMBER: 60/374,738  
 ; PRIOR FILING DATE: 2002-04-23  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 367  
 ; SOFTWARE: CuraseqList version 0.1  
 ; SEQ ID NO 6  
 ; LENGTH: 1120  
 ; TYPE: PR1  
 ; ORGANISM: Homo sapiens  
 ; US-10-262-839-6  
  
 Query Match 36.4%; Score 2224.5; DB 15; Length 1120;  
 Best Local Similarity 40.6%; Pred. No. 1.3e-187;  
 Matches 486; Conservative 199; Mismatches 397; Indels 115; Gaps 17;  
  
 QY 1 MELPFFVTHLFLPLVFLUTGLCSFNDLHHPRLLPFGPPEABFGYSVLQHVGGGORMMLVGA 60  
 DB 1 MDLPRLGLVAVALLSMPGFTDFTNMDTRKPRVPGTSRTAFFGTYTQQHDSGNKWLVLGA 60  
 QY 61 PWDGSGDRRGDVRCPVGAHNAACAHLGIDYQLGNSSHPAVNMHGLMSLLETDGCG 120  
 DB 61 PLETNGYQKTDVYKCPV---IHGNCNKLNG----- 89  
 QY 121 FMACAPLWSRACSSVFSSGICARVDASFPQGS LAPTAQRCTYMDVIVVLGDSNSIYP 180  
 DB 90 -----CQTMDIVVLGDSNSIYP 108  
 QY 181 WSEVQTFRLRLVGLKFLIDPEQIQVLGVQYGESPVHWSLGDPRTKBEVVRAAKNLSRREG 240  
 DB 109 WVEVQHFLINILKFKYIPGQIQGVVQYGEDVWHEFLNDVRSVKDVVEAASHIEQRGG 168  
 QY 241 RETKTAQIAIWAICTGFSQSHGRPEAARLLVVVDGSHDGEELPAALKAACEAGRVTRY 300  
 DB 169 TETRTAFGLFEARSEAFQK--GGRGAKKVMITVTDGSHSDPDLKVKIQOOSRDNVTRY 226  
 QY 301 GIAVLGHYLRORRDPSSFLREIRTIASDPDPFFNVTDAAALTDI VALGDRIFFLEGS 360  
 DB 227 AVAVLGYNNRGINPTFLNEIKYIASDPDDKHFNVDDEAALKDVALGDRIFFLEGT 286  
 QY 361 HAENESSFGLMSIQIGFSTRHLKDGILFGMVGAYDMGGSVLWLEGGHRLFPFRMALEDEF 420  
 DB 287 N-KNETSFGLMSQTFSSHVVEDGVLGAVGAYDWNGAVLKETSAGKVIPLRESYLKEF 345  
 QY 421 PPALQNHAAIYLGYSVSMILRGRRLLFLSGAPFRHRKVIAPOLKKGAVRVAQSLQGE 480  
 DB 346 PEEKKNHGAIYLGTVTSVSVSRQGRVYVAGAPFRNHTGKVLFTMHNRSRLTIHQMRGQ 405  
 QY 481 QIGSYFSGELCPDLTDRTDGTVDLLVAAAPMFLGPQKGTGRVYVYLVGQOSSLTLTQGLQ 540  
 DB 406 QIGSYFSGELTSVDIDGDTVDLLVGAAPMYFN-EGREGRKGYVYVEL-RQNRVYVNGTLK 463  
 QY 541 PEPP-QDARFGFAMGALPDINQDGFADVAVGAPLEDGHQALYLYHGTSQGVVRHPAQR 599  
 DB 464 DSHSYQNARFGSSIASVRDLNQDSYNDVVVGAPLEDNHAGAIYIIFHGFSGSLTKPKQRI 523  
 QY 600 AASMPHALSYFGRSVDGRDLDDGDLVDVAVGAQAALLSSRPVHLHTPSLEVTPOAI 659  
 DB 524 TASELATGLQYFGCSIHGQDLNEDGLIDLAVGALGNVILMSRPVQINASLHPEPSKI 583  
 QY 660 SVVQRCRGRGQAVCLTAALCPQVTSRTTPGRWDHQFYVMRFASLDEWTAGARAFDGS 719  
 DB 584 NIFHRCKSGRDRATCLAAFLCTPTPLAPHFQTTVGRIRYNATMDERRYPTRAHLDEGG 643  
 QY 720 QRLSPRLRLSVGNVTCQELHFLVLTDSYLPALVTFTFALDNTTKPGPVLINEGSPTSI 779  
 DB 644 DRETNRAVLLSSQELCERINFHLVLTADYVVPVTSVEYSLEDPH-GPMLDDGWPTTL 702  
 QY 780 QKLVPSKCGPDNECVTDILVQVNMDI-----RGRK-----APFV 817  
 ;



QY 774 GSPTSIOKLVFPKDCGPDNECVTDVLQVNMDIRGSRKAPFVVRGGRRKVLSTTLENR 833  
DB 793 SLFNSVHEYIPFAKDCGCKEKCISDLHLV-----ATTEKOLLIVRSQNDKFNVSITVKNT 848  
QY 834 KENAYNTSLGSIIFSRNLHLASLTQRESPIKVECAAPSAAHARLCSVGHVPFQCGAKVTEL 893  
DB 849 KDSAYNTRTVHVSFNLVFSGI-----EAIQKDSCE--ESNHNITCKVGYFPLRGEMVTFK 902  
QY 894 LEFEFSSLLSQVFGKLTASSDSLRNGTLQENTQTSAYIOYEPHLLSSSESTLHRYE 953  
DB 903 ILFOFNTSYLMENVTIYLSATSDSEEPETLSDNVNINISIPVKYEVGLQFVSSASEYHIS 962  
QY 954 VHPYGTLP-----VGPGEPEKTLTVQNLGCVVSGLIISALLPAVAHGN-----YFLSL 1004  
DB 963 IAAETVPEVINSTEDIGNEINIFYLIRKSGSPMPPELKLISFPNMTSGYPVLYPTGL 1022  
QY 1005 SQVITNASCIVQNLTEP-----PGPPVHPPEELQHTNRLNGSNTOCVVRCHLGQAKG 1058  
DB 1023 SS--SENANCRPHIFEDPFSINSKMTTSTDHLKRGTLDCNTCKEATTICNLTS--SDI 1079  
QY 1059 TEVSUGLRLVHNEFFERRAKFKSLTVVSTFELGTEGSLVQLTEASRWSLELLEVQOT-R 1117  
DB 1080 SQNVSL--ILWKPTFKSYFSSNLNLTIRGELSENAS-LVLSSSNQKRELAIQISKDGL 1136  
QY 1118 PILISLWILGSLVGLLLALLVFCVWKLGFFFAHKKIPBEEKREE 1163  
DB 1137 PGRVPLVILLSAFAGLLMLLILALWKIGTF---KRPLKKKMEK 1179

RESULT 13  
US-09-984-130-103  
; Sequence 103, Application US/09984130  
; Publication No. US2003005231A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 12 Human Secreted Proteins  
; FILE REFERENCE: P48992  
; CURRENT APPLICATION NUMBER: US/09/984,130  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR FILING DATE: 2000-10-30  
; PRIOR FILING DATE: 2000-10-30  
; PRIOR FILING DATE: 2001-04-18  
; PRIOR FILING DATE: 2000-04-19  
; PRIOR FILING DATE: 1999-10-27  
; PRIOR FILING DATE: 1998-10-28  
; NUMBER OF SEQ ID NOS: 149  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 103  
; LENGTH: 1151  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-984-130-103

Query Match 30.98; Score 1887; DB 10; Length 1151;  
Best Local Similarity 36.38; Pred. No. 1.4e-157;  
Matches 429; Conservative 226; Mismatches 454; Indels 74; Gaps 22;

QY 23 FNLDHHPRLFPFGPEAEFGYSLVQHVGGGQRMWLVGAPMDGPGSDRRGDYVRCVGAH 82  
DB 1 FNVDKNSMTFSGPVEDMFGYVQYENESGKWLIGSLVPLGQPKNKTGDKYKCPVGRGE 60  
QY 83 NAPCAKHLGDYQVGNSSHPAV-----NMHLGMSLLETDDGGGFMACAPLWSRACGSSVF 137  
DB 61 SLPCVKLDLP-----VNTSIPNVTVEKNMTFG--STLVTPNPGGFLACGPLYAYRCGHLHY 115  
QY 138 SSGICARVDASFOQGSGLAPTAQCPYMDVIVLDCGNSIYPWSEVQTFLLRLVGLKFI 197  
DB 116 TTGICSDVSPTFQVNVSIAP--VQECSTQLDIVIVLDCGNSIYPWSDVTAFLNDLLKRMDI 174

RESULT 14  
US-09-836-353A-103  
; Sequence 103, Application US/09836353A

QY 198 DPQIOQVLVOYGESPVHWSLGDFTKKEVVRRAAKNLSRRREGRETTTAQAIMVACTEGF 257  
DB 175 GPKQTOVGIVQYGENVTHEFNANKYSSTEEVLAACKIVQGGRTMTALGTDTPARKEAF 234  
QY 258 SOSHGGRPEAARLLLVVYDGSDEGEELPAALKACEAGRVTGVIAGLVGHVLRQRDPSS 317  
DB 235 TEARGARGVKVMVIVTDSHNDHRLKVKIQCEDENIQRFSIALIGSVNRNLSTEK 294  
QY 318 FLREIRTIASDPDERFFNFVNTDEAALTDIYDALGDRIFGLEGSHAENESSFGLMSQIGF 377  
DB 295 FVEEIKSIASEPTEKHEFFNVSDDELALVTIVKTIGERIFALEATADQSAASEMEMSQTF 354  
QY 378 STHRLKDGILFGMVAGYDWCSSVLWLEGHRLPPRMALEDEPPPALONHAYLYGYSVSS 437  
DB 355 SAHYSQDWMVLGAVGADWNGTVMQKASQIIIPRNTTFNVESTKQNEPLASYLYGTVNS 414  
QY 438 MLLRGGRRLFLSGAPRHRGKVIAFOLKKOGAVRVAQSGEOIGSYFSGSELCPDTR 497  
DB 415 ATASSGDVLTIAGOPRNYHTGQVIIYRM--EDGNIKILQILSGEIGSYFSGILTTTDDK 473  
QY 498 DGTDTLLVLAAPMLPGPQNKETGRVYVYLVGQOSLLTLOQTLP----- 541  
DB 474 DSNTDILLVGA PMVMTGEKEEQKVYVAL--NQTRFEYQMSLEPIKOTCCSSRQHNSCTT 532  
QY 542 ---EPPODARFGFAMGALPDINODGFADVAVGAPLEDHOGALYLYHGTQSGVRRPHPAQR 598  
DB 533 ENKNEPCGARFGTAIAVAKDLNLDGFNDIVIGAPLEDHOGAVVYIYHSGKTIIRKVAQR 592  
QY 599 IAAASPHALSYFGRSVDGRLDGLDVLVAVAGCAAGAAILLSSRPVHLTPSLVETPQA 658  
DB 593 IPSGDGLTKLFFGQS IHGEMDNLGDLTDTVTIGGLGAALFWSRDVAVVAVKVTWTFENK 652  
QY 659 ISVVQRDCRRRGOEAVCLTAALCFQVTSRTPGRMDHOFYMRFTASLDDEWAGARAAFDGS 718  
DB 653 VNIQKNCHMEGKETVCINATVCFEVLKSKEDTIYADLQYRVTLDSLRQISRSFSGT 712  
QY 719 GQRLSPRRLSLVGNVTCEQLHFVLDTSYLRPALVTFTFALDNTTKP--GPVLNEGSP 776  
DB 713 QERKQVR--NITVRKSECTKHSFVMLDKHDFQDSVRITLDF--NLTDPENGPVLDDSLP 767  
QY 777 TSJQKLVPSKDCGPDNECVTDVLQVNMDIRGSRKAPFVVRGGRRKVLSTTLENKEN 836  
DB 768 NSVHEYIPFAKDCGCKEKCISDLHLV-----ATTEKOLLIVRSQNDKFNVSITVKNTKOS 823  
QY 837 AYNTSLSIIFSRNLHLASLTQRESPIKVECAAPSAAHARLCSVGHVPFQCGAKVTFLLF 896  
DB 824 AYNTRTIVHVSFNLVFSGI-----EAIQKDSCE--ESNHNITCKVGYFPLRGEMVTFKILF 877  
QY 897 EFSCSSLLSQVFGKLTASSDSLRNGTLQENTQTSAYIOYEPHLLSSSESTLHRYE VHP 956  
DB 878 QFNTSYLMENVTIYLSATSDSEEPETLSDNVNINISIPVKYEVGLQFVSSASEYHISAA 937  
QY 957 YGTLP-----VGPGEPEKTLTVQNLGCVVSGLIISALLPAVAHGN-----YFLSLSQV 1007  
DB 938 NETVPEVINSTEDIGNEINIFYLIRKSGSPMPPELKLISFPNMTSGYPVLYPTGLSS-- 996  
QY 1008 ITNASCIVQNLTEP-----PGPPVHPPEELQHTNRLNGSNTOCVVRCHLGQAKGTEV 1061  
DB 997 -SENANCRPHIFEDPFSINSKMTTSTDHLKRGTLDCNTCKEATTICNLTS--SDISQV 1054  
QY 1062 SVGLLRLVHNEFFERRAKFKSLTVVSTFELGTEGSLVQLTEASRWSLELLEVQOT--RPIL 1120  
DB 1055 NVSL--ILWKPTFKSYFSSNLNLTIRGELSENAS-LVLSSSNQKRELAIQISKDGLPGR 1111  
QY 1121 ISLWILGSLVGLLLALLVFCVWKLGFFFAHKKIPBEEKREE 1163  
DB 1112 VPLVILLSAFAGLLMLLILALWKIGTF---KRPLKKKMEK 1151



```
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 103
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-353A-103

Query Match      30.8%; Score 1887; DB 10; Length 1151;
Best Local Similarity 36.3%; Pred. No. 1.4e-157;
Matches 429; Conservative 226; Mismatches 454; Indels 74; Gaps 22;

Qy 23 FNLDEHHPRLFPGPPEAEFGYSLQHVGGGQRMWLVGAPWDPGSGDRRGDVYRCPVCGAH 82
Db 1 FNVVDKNSMTFSGPVEDMFGYVQYENEGKWLIGSLVGPQPKNTGDKYKCPVGRGE 60
Qy 83 NAPCAKGLHDYQLGNSHPAV-----NMHLGMSLLETDDGGGPMACAPWRSRACGSSVF 137
Db 61 SLPCVKLDLP-----VNTSIPNVTEVKENMTFG-STLVTPNPGGFLACGPLYAVRCGHLHY 115
Qy 138 SSGICARVDASFOQPGSLAPTAQRCPYMDVIVLDGNSNITYPMSEVQTFRLRLVGLFI 197
Db 116 TTGTICSDVSPFTQVQVNSIAP-VQECSTQLDIVLDGNSNITYPMDSVTAFLNDLKKRMDI 174
Qy 198 DPEQIQVGLVQYQVGGSPVHMSLGRFRTKEEVRAAKNLSRREGRETKTAQAIMVACTEGF 257
Db 175 GPQKTQVGIQYQVGENVTHEFNLNKYSSTEEVLVAAKIVQRGGRQWTALGTDTARKEAF 234
Qy 258 SQSHGGRPEARLLVVVDGESHDPGEELPAALKAACEAGRVTRYGIAGVILGHYLRQRDPSS 317
Db 235 TEARGARGVKKVMVIVTDGESHDNHRLKKVIOQCEDENIQRFSAIILGSYNRGNLSTEK 294
Qy 318 FLRIRTIADPDPRFPFNTDEAALTDIVDALGDRIFGLEGSHAENESSFGLMSQIGF 377
Db 295 FVEIKSIASEPTSKHFFNVSDLEALVTIVKTLGERIPALEATADQSAASFEMESQTGP 354
Qy 378 STHRLKDGILFGMYGAYDWGSGVLWLGGRHLRFLPPRMALEDEFPFALQNHAAVILGYSVSS 437
Db 355 SAHYSQDWVMLGAVGADWNGTVVMQKASQIIIPNTTFNVESTKKNEPLASVLYGVNS 414
Qy 438 MLLRGGRRLFLSGAPRFRHKGKVIJAFOLKXGAVRVAQSLQGEIGSYFSGELCPDLTDR 497
Db 415 ATASSGDVLYTAGQPRYNHTQVLIYRM-EDGNIKILQTLGSGIGSYFSGSILTTDIDK 473
Qy 498 DGTDDVLLVAAPMLFQNKETGRVYVYLVVQCSLLTLQGLTP----- 541
Db 474 DSNTDILLVGAPMYMGTEKEEGKVYVYAL-NQTRFYQMSLEPIKQTCSSSRQHNSCTT 532
Qy 542 ---BPPQDARFGFANGALPOLNQDGFADVAVGAPLEDGCHQALYLYHGTQSGVPHPAQR 598
Db 533 ENKNEPCARGFTALAAVKOLNLDGFNDIVIGAPLEDHGHGAVYIYHSGKTIKREYAQR 592
Qy 599 IAAASMPHALSYFGRSVDGRDLDDGDLVDVAVAGQAAILLSRPIVHLTPSLEVTPQA 658
Db 593 IPSGGDGKTLKFFQSGIHGMDLNGDGLTDVTIGGLGAALFWSRDAVAVKVTNNFEPNK 652
Qy 659 ISVVQRCCRGRGQAVCLTAALCFQVTSRTSGRDWHQFYMRFTASLDWNTAGARAPDGS 718
Db 653 VNIOKNCHMEGKETVCINATVCPEVKLUKSKEDTIIYADLUQYRVTLDSLRQISRSFSGT 712

; Publication No. US10741600
; Sequence 1161, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1161
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-741-600-1161

Query Match      30.8%; Score 1880; DB 17; Length 1177;
Best Local Similarity 36.3%; Pred. No. 6e-157;
Matches 430; Conservative 225; Mismatches 455; Indels 76; Gaps 23;

Qy 20 CSDPNDLDEHHPRLFPGPPEAEFGYSLQHVGGGQRMWLVGAPWDPGSGDRRGDVYRCPVG 79
Db 26 CVSNFVNDKNSMTFSGPVEDMFGYVQYENEGKWLIGSLVGPQPKNTGDKYKCPVG 85
Qy 80 GAINAPCAKGLHDYQLGNSHPAV-----NMHLGMSLLETDDGGGPMACAPWRSRACGS 134
Db 86 RGSLLPCVKLDLP-----VNTSIPNVTEVKENMTFG-STLVTPNPGGFLACGPLYAVRCGH 140
Qy 135 SVFSSGICARVDASFOQPGSLAPTAQRCPYMDVIVLDGNSNITYPMSEVQTFRLRLVGLK 194
Db 141 LHYTTGTICSDVSPFTQVQVNSIAP-VQECSTQLDIVLDGNSNITYPMDSVTAFLNDLLE 199
Qy 195 LFTDPEQIQVGLVQYQVGGSPVHMSLGRFRTKEEVRAAKNLSRREGRETKTAQAIMVACT 254
Db 200 MDGPKQTQVGIQYQVGENVTHEFNLNKYSSTEEVLVAAKIVQRGGRQWTALGTIDTARK 259
Qy 255 EGFSSQSHGGRPEARLLVVVDGESHDPGEELPAALKAACEAGRVTRYGIAGVILGHYLRQRD 314
```

Db 260 EATEARGARGVKVWVITDGHSHNHLKKVIOQCEDENIQRFSAIALGSYNRNLS 319  
Qy 315 PSSFLREIRTIASDPDRFFNTDEAALTDIVDALGDRIFGLEGSHAENESSFGLEMSQ 374  
Db 320 TEKFVEIKSIASEPTKEHFNVSDELALVTIVKTGLGERIFALEATADQSAASEMEMSQ 379  
Qy 375 IGFSTHRLKDGILFGMVGAYDWGVSUWLGHHRLFPFRMALEDEPPPALQNHAAIYLGYS 434  
Db 380 TGFSAHYSQDWMLGAVGAYDWNGTVVMQKASQIIIPRNTTFNVESTKKNPLASYLGYT 439  
Qy 435 VSSMLLRGRRFLSGAPRFRHGRKVIATFOLKDDGAVRVAQSLQGOIGSYFGSELCPD 494  
Db 440 VNSATASSGVDVLIAGQVRNHTGQVYIYRM-EDGNIKILQILSGEIGSYFGSILTTD 498  
Qy 495 TDRDGTDDVLLVAPMFLGPNKGTGRVYVYVVGQSSLLTLQGTLOP----- 541  
Db 499 IDKDSNTDILLVGAFTMGTEKEBQGVYVYAL-NQTRFEYQMSLEPIKQTCSSRQHNS 557  
Qy 542 -----EPPQDARFGFAMGALPDINQDGFADVAVGAPLEDHOGALYLYHGTSGVVRPHP 595  
Db 558 CTTENKNEPCGARFGTAIAAVKOLNLDGFNDIVIGAPLEDHGGAVYIYHSGKTKIRKEY 617  
Qy 596 AQRIAAASMPHALSYFGRSVDGLDLDGDDLDVAVCAQGAAILSSRPVHLTPSLEVT 655  
Db 618 AQRIPSGDGTUKFPQGSIHGENDLNGDGLTDVTICGLGGAALFWSRDAVAVKVTWNE 677  
Qy 656 POAISVVQRDCRRRGQAVCLTALCPQVTSRTFGRWDHOFYRFTASLDEBTAGAPAT 677  
Db 678 PNKVNIOKKCHMEGKETVCINATCFDVKLKSKEDTIEADLQYRVT 677  
Qy 716 DSGQRLSPRLRLSVGNVTCEQLHFHVLDSYLRPVALTVTFALDNTTKP--GPVLNE 773  
Db 738 SGTQERKQVR--NITVRKSECTKHSFYML--HDFQDSVRITLDF--NLDPENGPVLDD 790  
Qy 774 GSPTSIOQLVPSKCGPDNECVTDVLOVNMDIRGSRKAPFVVRGGRKVLVSTTLNR 833  
Db 791 SLPSNVHEYIPFAKDCGCKEKCSIDLHV-----ATTEKOLLIVRSQNDKFNVSILTAKNT 846  
Qy 834 KENAYNTSLIIFSRNLHLASLTPORESPIKVECAAPSAAHARLCSVGHVPVFTGAKVTF 893  
Db 847 KDSAYNTRTIVHYSPLVPSGI----EAIQKDS--ESNHNITCKVGYPLRGRGEMVTFK 900  
Qy 894 LEFEFCSSLLSQVFGKLTASSDSLERNGTLQENTATQTSAYIOYEPHLLFSSSESTLHRYE 953  
Db 901 ILFOFNTSYLMENVTVILSATSDSEEPETLSDNVVNISIPVKYEVGLQFVSSASEYHIS 960  
Qy 954 VHPYGTLP-----VGPGEFKTLRVQNLGCYVVSGLIISALLPAVAHGN---YFLSL 1004  
Db 961 IAAETVPEVINSTEDIGNEINIPYLIRKSGFPMPELKLISFPNMTSNGYPVLYPTGL 1020  
Qy 1005 SQVITNNASCIVQNLTEP-----PGPPVHPEELQHTNRLNGSNTQCQVVRCHLQGLAKG 1058  
Db 1021 SS--SENANCRPHIFEDFPFINSCKNMTTDLHKGRTILDNCCKTATTCNLTS-SDI 1077  
Qy 1059 TEVSVGLLRLVHNEFFFRAXFKSLTVVSTFELGTEGSLVQLTEASRWSLSLEVQOT-R 1117  
Db 1078 SQNVSL--ILWKPFTIKSVFSSNLRTIRGELRSENAS-LVLSSSNQKRELAIOISKDGL 1134  
Qy 1118 PILISLWILGVLGGLLALLVFLWKLGFPAHKKIPEEKREE 1163  
Db 1135 PGRVPLWILLSAFAGULLLMLLALWKGIF---KRPLKKQMEK 1177

Search completed: April 6, 2005, 12:53:00  
Job time : 151.761 secs

**This Page Intend (empty)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:52 ; Search time 49.8054 Seconds  
(without alignments)  
1749.117 Million cell updates/sec

Title: US-09-647-544-2  
Perfect score: 6106  
Sequence: 1 MELPFVTHLFLPLVFLTGLC.....GFFAHKKIPBEKKREKLEQ 1167

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A-COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B-COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A-COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B-COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS-COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2429.5	39.8	1217	4	US-09-949-016-7892
2	1894.5	31.0	1180	4	US-09-000-004A-2
3	1784	29.2	1183	4	US-09-532-310B-5
4	1756	28.8	1181	4	US-09-000-004A-4
5	1756	28.8	1181	4	US-09-949-016-6189
6	1756	28.8	1195	4	US-09-949-016-10747
7	1728	28.3	1183	4	US-09-532-310B-6
8	1131.5	18.5	1161	1	US-08-173-497-2
9	1131.5	18.5	1161	1	US-08-286-889-2
10	1131.5	18.5	1161	1	US-08-485-618-2
11	1131.5	18.5	1161	1	US-08-362-652-2
12	1131.5	18.5	1161	2	US-08-605-672-2
13	1131.5	18.5	1161	2	US-08-482-293A-2
14	1131.5	18.5	1161	2	US-08-943-363-2
15	1131.5	18.5	1161	3	US-09-193-043-2
16	1131.5	18.5	1161	4	US-09-688-307A-2
17	1131.5	18.5	1161	4	US-09-350-259-2
18	1126	18.4	1161	1	US-08-485-618-99
19	1126	18.4	1161	2	US-08-605-672-99
20	1126	18.4	1161	2	US-08-482-293A-99
21	1126	18.4	1161	2	US-08-943-363-99
22	1126	18.4	1161	3	US-09-193-043-99
23	1126	18.4	1161	4	US-09-688-307A-99
24	1126	18.4	1161	4	US-09-350-259-99
25	1124.5	18.4	1161	3	US-09-193-043-55
26	1124.5	18.4	1161	4	US-09-688-307A-55
27	1124.5	18.4	1161	4	US-09-350-259-55

28	1118.5	18.3	1161	1	US-08-485-618-55	Sequence 55, Appl
29	1118.5	18.3	1161	1	US-08-362-652-55	Sequence 55, Appl
30	1118.5	18.3	1161	2	US-08-605-672-55	Sequence 55, Appl
31	1118.5	18.3	1161	2	US-08-482-293A-55	Sequence 55, Appl
32	1118.5	18.3	1161	2	US-08-943-363-55	Sequence 55, Appl
33	1109	18.2	1161	1	US-08-485-618-53	Sequence 53, Appl
34	1109	18.2	1161	1	US-08-362-652-53	Sequence 53, Appl
35	1109	18.2	1161	2	US-08-605-672-53	Sequence 53, Appl
36	1109	18.2	1161	2	US-08-482-293A-53	Sequence 53, Appl
37	1109	18.2	1161	2	US-08-943-363-53	Sequence 53, Appl
38	1109	18.2	1161	3	US-09-193-043-53	Sequence 53, Appl
39	1109	18.2	1161	4	US-09-688-307A-53	Sequence 53, Appl
40	1109	18.2	1161	4	US-09-350-259-53	Sequence 53, Appl
41	1108.5	18.2	1151	1	US-08-286-889-37	Sequence 37, Appl
42	1108.5	18.2	1151	1	US-08-485-618-37	Sequence 37, Appl
43	1108.5	18.2	1151	1	US-08-362-652-37	Sequence 37, Appl
44	1108.5	18.2	1151	2	US-08-605-672-37	Sequence 37, Appl
45	1108.5	18.2	1151	2	US-08-482-293A-37	Sequence 37, Appl

ALIGNMENTS

RESULT 1

US-09-949-016-7892

; Sequence 7892, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7892

; LENGTH: 1217

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-7892

Query Match 39.8%; Score 2429.5; DB 4; Length 1217;

Best Local Similarity 43.1%; Pred. No. 1.6e-225;

Matches 516; Conservative 213; Mismatches 422; Indels 45; Gaps 16;

Qy 1 MELPFVTHLFLPLVFLTGLCSPFNLDHHPRLFPGPPEAEFGYSVLQHVGGQRMWLVGA 60

Db 30 MDLPRGLVVAWALSFWFGFTDTFMDTRKPRVIFGSRTPFGYTVQQHDSGNKWLWVGA 89

Qy 61 PWDGPGDRRGGVYRCPVGGAHNAAPCAKHGLDYLQGNSSHPAVNMHLGMSLLETDCGG 120

Db 90 PLENTGYQKTGDVYKCPV---IHGNCITKLNLGRVTLNSVSRKDNMRGLSLATNPXDNS 146

Qy 121 FMACAPIWSRACGSSVSSFGICARVDASFQGGSLATPACQRCPTMYMDVVDGNSNIYP 180

Db 147 FLACSPLSWHECGSSYTTGMSRVNSFRFSKTPAVALQRCQTYMDIVIVLDGNSNIYP 206

Qy 181 WSEVQTLRLRLVGLFIDPEQIQVGLVQYGSPPHWSLGDFTKEEVVRAAKNLSRREG 240

Db 207 WVEVQHFLINILKFKYIGPGQIQGVVQYQGVGVVHEFLINDYRSVKDVVEAASHIEQGG 266

Qy 241 RETKTAQAIMVACTEGFSQSHGRPEARLAVVVVTDGSHDGEELPAALKAACEAGRVTY 300

Db 267 TETRTAFGIEPAREAFQK--GGRKGAKVMIVITDGESHDSPLDKVIQQSRDNTVRY 324

Qy 301 GIAVLGHLRQRDPSSFLIRTIASDPDRFFNVNTDEAALTDIVDALGDRIFGLEGS 360

Db	325	AVAVLGYNRRGINPETFLNEIKYIASPDDBKHFNVTDEALKXDIDVALGDRIFFSLEGT	384
Qy	361	HAENESSFGLEMSQIGFSTHRLKDGILFGMYGAVDWGSGVJLWEGGHRFLFPPRMALEDEF	420
Db	385	N-KNETSFGLEMSQTGFSSHVEDGVLLGAVGYDMNGAVLKETSAGKVIPLRESYLKEP	443
Qy	421	PPALQNHAAIYLGYSVSSMLRGGRRLFLGSGAPRFRHROKVIAPQLKDGAVRVQAOSLOGE	480
Db	444	PEELKNHGAYLGIVTYSVVSSRQGRVYVAGAPRPNHTKGVILFTWMHNRNSTIHOAMRGQ	503
Qy	481	QIGSYFSGSELCPDLDRDGTVDLLVAAPMFLGPNKETGRTGVYVYLVGQOSLLTLTQGTLO	540
Db	504	QIGSYFSGEITSVDLDGGVTDVLLVGAPMYFN-EGRERGKVYVYVEL-RQNRFFVYNGTLK	561
Qy	541	PEPP-QDARFQFAMGAPDLNQDGFADVAVGAPLEDHGQALYLVHGTSQGVRRHPHQAORI	599
Db	562	DSHSYQONARFGSSIASVRDLNQDSVNDVVGAPLEDNHAGAIYIFHGFPGSILKTPKQRI	621
Qy	600	AAASMPHIALSYFGRSVQGRDLDDGDDLVDVAVGAGQAAILSSRPIVHLTPSLEVTPQAI	659
Db	622	TASSELATGLQYFGGSIHQGLDNLNEDGLTDLAVGALGNAVILWSRPPVQINASLHPEPSKI	681
Qy	660	SVVORDCRRRQBAVCLTAALCFQVTSRTPGRWDHQFYMRFTASLDEWTAGARAAPGSG	719
Db	682	NIFHRDCKRSRDATCLAAFLCFTPIFLAPHFQTTVGIRYNATMDERRYTPRAHLDDEGG	741
Qy	720	QRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALTVTFALDNTTKPGPVLNEGSPTSI	779
Db	742	DRFTNRAVLSGGQELCERINFHVLDTADYVKPVTFSVEYSLEDPDH-GPMLDDGWPTTL	800
Qy	780	QKLVPPSKDCGPDNECVTDVLQVNMDOI-----RGRK-----APFV	817
Db	801	RVSYPFFWNGCNEDEBHCVPDVLVDARSDLPTAMEYQCVLRKPAQDCSAYTLTSFDTTVII	860
Qy	818	RGGRKVLVSTTLNKRKENAYNTSIIIPSRNLHLASLTPQRESPIKVECAAPS--AHAR	875
Db	861	ESTQRVAVEATLENRGENAYSTVLNISQSANLOFASLIQKEDSGSTECVNEERRLOKQ	920
Qy	876	LCSYGHVPFQGAQVTFLLBEPFSCSLLSVOFGKLTASSDSLBERNGTLOENTACTSAYI	935
Db	921	VCNVSYPPFRKAKAVAPLDFEFSKSIPLHLHLEIELAAGSDSNERDSTKEDNVAPLRPHL	980
Qy	936	QYEPHLLFSSBESTLHRVBYHYGTLP--VGCGPEFKTTLRVONLGYCVVSGLIISALLPA	993
Db	981	KYEADVLFTRSSSLSHYEVKLNLSLERYDGGPPFSCIFRIQNLGLLFLPIHGWMKITTPI	1041
Qy	994	VAHGGNYFLSLSQVITN--NASC-IVQNLTEPPGPPVHPELOHTNRLNGSNTQCVVRC	1050
Db	1041	ATRSGRNLLKLDRDLTTEANTSCNIWGNSTBYRTPVE-EDLRAPQLNHSNDSOVVSINC	1098
Qy	1051	HLGQAKGTEVSVGLLRVLVHNEFPFRRAKFSLTVVSTFELTEBGSVLQLTASRWSSSL	1110
Db	1100	NI-RLVPNOEINFHLLGNLRSUKALKYSKMIWMAALQRQHSPPIFREEDPSRQIV	1158
Qy	1111	LEVQVTPRILISLWILGSVLGGLLALLVFCVLWKGFFPAHKKIPBEEKREKLE	1166
Db	1159	FEISKQBDQWPIWIIIVGSTVGLLALLVWLKGLFFFSAR----RRREPGID	1210

```

; PRIOR APPLICATION NUMBER: US 60/001,387
; PRIOR FILING DATE: 1995-07-21
; PRIOR APPLICATION NUMBER: US 60/001,861
; PRIOR FILING DATE: 1995-08-03
; PRIOR APPLICATION NUMBER: US 60/016,700
; PRIOR FILING DATE: 1996-05-02
; PRIOR APPLICATION NUMBER: PCT/US96/12067
; PRIOR FILING DATE: 1996-07-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1180
; TYPE: PRT
; ORGANISM: Rattus
US-09-000-004A-2

Query March 31.08; Score 1894.5; DB 4; Length 1180;
Best Local Similarity 36.5%; Pred. No. 1.1e-173;
Matches 436; Conservative 221; Mismatches 461; Indels 77; Gaps 24;

QY 13 LVFLTGTCSPFNLDDEHHPRLFPGPPEAEFGYSVLQVHGQGMWLVGAPMDGPSGDRRGD 72
DB 19 LTWILGFCVFENVVDVKNMSFGSFGVEDMFYTVQQYENEBGKWLVTGSLVUGQPKARTGD 78
QY 73 VYRCPVGGAHNAPCAKGHLGDYOLGNSHPAV-----NMHLGMSLLLETDGCGFMACAPL 127
DB 79 VYKCPVGRERAMPCKVLDLP---VNTSIPNVTEIKENMTFG-STLVTPNPGFLACGPL 133
QY 128 WSRACGSSVFSSGICARVDASFOQGS LAPTAQRCTPYMDVIVL DGSNSIYPWSEVQTF 187
DB 134 YAVRCGHLYHTTGICSDSVPTFQVNSFAP-VQECSTQLDITVILVDGSNSIYPWESVIAF 192
QY 188 LRLRLVGKGLFDPEIQVLQVQYGESPVHSEISGDFTKBEVRAAKNLSRREGRETAKTQ 247
DB 193 LNDLRLKMLGPTQTVGIQVQYENVTHEFNLNKYSITSEELVVAANKIGQGGLQMTAL 252
QY 248 AIMVACTEFSQSHGGRPEARLLVVVTDGESHGDBELPAALKACEAGRVTRYGIAVLGH 307
DB 253 GIDTARKEAFTEAGARRGVKKVMVITDGE SHDNVRLKQVIQDCEDENIQRFSIALGH 312
QY 308 YLRQRDPDSFLRHRTIASDPBERFPFNVTDBAALTDIVDALGDRIFGLEGSHAENESS 367
DB 313 YNRGNLSTEFKVFBEIKSIASEPTEKHPFNVSDELALVTIVKALGERIFALEATAQSAAS 372
QY 368 FGLEMSOIGFSTRHLKDGII LFGMVGAYDMCGSVLWLEGGHRLPPRMALDEFPPLQNH 427
DB 373 FEMSMSTGFSAHYSQDVMGLGAVGYDMNGTVVMQKANOQMVIPHNTTPTQTE--PARONE 430
QY 428 --AAYLGYSVSSMLLRGGRFLTSGAPFRFHRGKVIAFLQKKGAVRVQAQSLQGEQIGSY 485
DB 431 PLASYLGTVNSATIPGD-VLYTAGQPRYNHTGOVVIYKM-EDGNINILQTLGGEQIGSY 488
QY 486 FGSLELCPDLTRDGTDTDLLVAAPMFLGPONKETGRVYVYLVGQQSLLTLQGTQLP---- 541
DB 489 FGSVLTATIDKDSYTDLLLVGAPMYMTGTEKEBQGVYVYAV-NQTRFEBYQMSLEPIROT 547
QY 542 -----EPPQDARFGFANGALPDLNODGFADVAGAPLEDHQCALVLYHG 586
DB 548 CCSGLKDNSTCKENKNEPCGARGFTAAVKDLNVDFGNDVVIAGPLEDDHAGAVIYHG 607
QY 587 TQSGVRPHPAQRIATAASMPHALSYFGRSDVGRDLDDGDDLVDVAVAGQAAGAILLSRPV 646
DB 608 SGKTI RAYAQRI PSGGDGKTLAFQGS IHGEMDLNGDGLTDVTI GGLGGAALFWARDVA 667
QY 647 HLTPSLBVTPOAISVQVORDCRRRQGBAVCLTAAALCFQVTSRTPGRWDHOFYMRFTASLDE 706
DB 668 VVKVTMNFENKVINIQKNCRVBGKETVCINATMCFHVKLKSKEDSIIYEADLQYRVTLDS 727
QY 707 WTAGARAAFGSGORLSPRRLRLSVGNVTCEQLHFFHVLDTSDYLRPVALTVTFTALONTTK 766
DB 728 LRQISRSFFGQERKIQR--NITVRESECI RSHSFYMLDKDHQDQDSVRVTLDF---NLTD 782
QY 767 P--GPVLNIGSPSTSIQKLVFPFSKQCPDNCEVTDLVLVQNMMDIRGSRKAPFVVVRGRRKV 824

```

Db 783 PENCGLDDALPNSVHEHIPPACOGKNCERCISDLTLNVST-----TEKSLIIVKSQDKF 838  
Qy 825 LVSTTLNKENAYNTSLSIIFSRNLHLASTP-QRESPIKVECAAPSAHARLCSVGHVP 883  
Db 839 NVSLTVKNKGDANTRTVQHSFNLIPIGSEIEIKDS-----CESNQNTCRVGVYP 891  
Qy 884 FOTGAKVTFLEPFSCSSLLSQVFGKLTASSDLENGTLQENTAOISAVIQEPHLLP 943  
Db 892 LRAGETVTFKIIFOFNTSHLSNAIIHLSATSDSEEPLESLENDNEVNISIPVKYEVGLQF 951  
Qy 944 SSESTLHRYEHPYGTLP-----VGPQPEKTLRVQNLCCYVVSGLIISALLP-AVAH 996  
Db 952 YSSASEHHISVAANETIPEFNTSTEDIGNEINVPYTIKRGHPMPPELOLSIFPNLTAD 1011  
Qy 997 GGNVFLSQVITNVASCIVQNLTEPPG-----PPVHPELQHTNRLNGSNTOCQVVR 1050  
Db 1012 GYPVLYPIGWSSDNVNCPRSLDPFGINGCKWMTISKSEVLKRGTTIDCCSTCGVATI 1071  
Qy 1051 HLGOLAKG-TEVSVGLRLVHNEFFRRAKFKSLTVSTFELGTGEGSVLQLTEASRWSES 1109  
Db 1072 TCSLLPSDLQVNVSL--LLWKPTFIRAHFSLNLTLRGELKSENS--LTLSSNRKREL 1128  
Qy 1110 LLEVQOT-RPTLSILWILIGSVLGLLLALLVFLCWLKLGFAHKKIPPEEKREE 1163  
Db 1129 AIQSKDGLPQVPLWILLSAFAGLLMLLLIALLWKIGFF----KRLPKKKMEK 1180

## RESULT 3

US-09-532-310B-5

; Sequence 5, Application US/09532310B

; Patent No. 6596276

; GENERAL INFORMATION:

; APPLICANT: Senger, Donald R

; Detmar, Michael

; Claffey, Kevin P

; TITLE OF INVENTION: Method for inhibiting tumor

; angiogenesis in a living subject

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: David Prashker, Esq.

; STREET: P.O. Box 5387

; CITY: Magnolia

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 01930

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage

; COMPUTER: Dell PC

; OPERATING SYSTEM: MS DOS

; SOFTWARE: Microsoft Word version 97

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/532,310B

; FILING DATE: 22-Mar-2000

; CLASSIFICATION: Unknown

; ATTORNEY/AGENT INFORMATION:

; NAME: David Prashker, Esq.

; REGISTRATION NUMBER: 29,693

; REFERENCE/DOCKET NUMBER: BIS-036

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (978) 525-3794

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1183 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-532-310B-5

Query Match 29.2%; Score 1784; DB 4; Length 1183;

Best Local Similarity 35.2%; Pred. No. 5.7e-163;

Matches 421; Conservative 222; Mismatches 474; Indels 80; Gaps 24;  
Qy 13 LVFLTGLCSFNLDEHHPRLFPGPPEABFGYSVLQHVGGQRMMLVGAHPDGPDSGDRRGD 72  
Db 21 LTVILGFCVFNVDVKMSMTFSGPVEDMFGYTVQOYENEEGKWVLIGSPLVGQPKNRTGD 80  
Qy 73 VYRCPVCGAHNACAKGHLDGYQLGNSSHPAV-----NMHLGMSLLTDCDGGGFWACAPL 127  
Db 81 VYKCPVGRGESLPCVKLDLP-----VNTSIPNVTVEKNMTFG-STLVNPNNGGLACGPL 135  
Qy 128 WSRACSSSVFSGICARVDASFPQGS LAPTAQRCPTYMDVIVLDGNSNIYPPSEVQTF 187  
Db 136 YAYRCGHLHYTGLCSDVSPTFQVNSIAP-VQECSTXLTDIVIVLDGNSNIYPPMSV-TA 193  
Qy 188 LRLVLGKLTDPQIOVLQVYGESPVHWSLGDFTKKEEVRRAAKNLSRREGRETAKQ 247  
Db 194 LNDLLKEMDGPCKXTXGVIKXGENVTHEFNLNKYSSTEELVAAKKIVXRGGRXTWTAL 253  
Qy 248 AIMVACTEGFSQSHGGRPEAARLLVVVTDGESHDEELPAALKACAGRVTRYGIAVLGH 307  
Db 254 GTDTARKEATEARGARRGVKKVMVIVTDEHDXNHLKKVIGDCEDENIXRFIAILGS 313  
Qy 308 YLRRQDPSPFLRIRTIASDPDERFFFNVTDEAALTDIVDALGDRIFGLEGSHAENESS 367  
Db 314 YNRGNLSTEXFVEIKSIASEPTEKSFNVSDELALVTIKTGLGERIFALEATADXSAAS 373  
Qy 368 FGLMSOIGFSTHRLKDGILFGVMGVADWGSVUWLEGGHRLFPPRMALEDEFPALQNH 427  
Db 374 FEMEMSGTGFSAHYSQDMVWLGAAYDWMGTVVNQKASQIIIPRNTTFNVESTKNEPL 433  
Qy 428 AAYLGYSVSMMLRGGRRLFLSGAPRFRHRGKVTAFLQKKDGA VRAQSGOGEIGSYFG 487  
Db 434 ASYLGTVNSATASSGDVLYIAGPRYNTHTGQVYIYRM-BEGNIKIQTLSGEXIGSYFG 492  
Qy 488 SELCPLDTRDGTDLVLLVAAPMFLGPQNKETGRVYVYLVGQQSLTLTQGTLP-EP--- 543  
Db 493 SILTTTIDIKDSNTNILLVGAPMYMGTEKEGGKVYVYAL-NQTRFVQNSLAPMEPIKQ 551  
Qy 544 -----PQDARFGAMGALPDNLQDGFADVAVGAPLEDHQGALYLYH 585  
Db 552 TCCSSRQHNSCTTENKNEPCGARPCTAAVKOLNLDGFNDIVIGAPLEDHGGAVYIYH 611  
Qy 586 GTQSGVYPHPAQRIAAASMPHALSYFGSDVGRDLDDGLDVAVAGQAAILLSRPI 645  
Db 612 GSGKTIKEXYXIRIPSGGDKTLKFFGOSIHGMDLNGDLTDVTICGLGAALFWSRDV 671  
Qy 646 VHLTPSLEVTPOAISVVQDRCRRRQBAVCLTAALCFQVTSRTPGRWDHOFYMERFTASLD 705  
Db 672 AVVKVTNFPFNKVNIOKKNCHMEGKETVCINATVCEVULKSKEDITYEADLQYRVTL 731  
Qy 706 EWTAGAAAFDGGQRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRLPVALTPTFALDNTT 765  
Db 732 SLRQISRSFFSGTQERKVQR-NITVRKSECTSHSFYMLDKHDPQDSVRIITLDF---NLT 786  
Qy 766 KP--GPVLNEGSPTSIOKLVPFSKDCPDNECVTDVLOVNMDIRGSKAPFVVRGGRK 823  
Db 787 DPENGVLDDSLPNSVHEYIPFAKDCNCKEKCISDLSHV-----ATTEKDLIIVRSONDK 842  
Qy 824 VLVSITLENKENAYNTSLSIIFSRNLHLASTLPQRESPIKVECAAPSAHARLCSVGHVP 883  
Db 843 FNVSLTVKNTKDSAYNTRTIVHYSPNLVFSGI-----EAIQKDC--ESNHNITCKVGYFP 896  
Qy 884 FQTGAKVTFLEPFSCSSLLSQVFGKLTASSDLENGTLQENTAOISAVIQEPHLLP 943  
Db 897 LRRGEMVTFKILFOFNTSYLMENVTIYLSATSDSEEPETLSNDNVNISIIPVKYEVGLQF 956  
Qy 944 SSESTLHRYEHPYGTLP-----VGPQPEKTLRVQNLCCYVVSGLIISALLPAVAHG 997  
Db 957 YSSASEYHISIAANETVPEVINSTEDIGNEINIFYLIRKSGSFPMPPELKLKLSISFPNMTSN 1016  
Qy 998 GNYFL---SLSQVITNNAACIVQNLTEP-----PGPPVHPEELQHTNRLNGSNTOCQV 1048  
Db 1017 GYPVLVPTGLSS--SENANCRPHIFEDFPFSINSKKWTTSTDLKRGRTILDCTCKPATI 1074





```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6189
; LENGTH: 1181
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6189

Query Match      28.8%; Score 1756; DB 4; Length 1181;
Best Local Similarity 34.3%; Pred. No. 3e-160;
Matches 418; Conservative 219; Mismatches 455; Indels 126; Gaps 31;

Qy 11 LPLVFLTGL-----CSPFNLDHHPRLFPGPPEAFSGYVLQHVGGQRMVLVGPWD 63
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 11 LPLLLVLSOGILNCLLAYNVGLPEAKIFSGPSEQFGYAVQVQFINPKGNWLLVGPWS 70
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 64 GPSGDRGDVYRCPVGGAHNAPCAKHL-GDYQGNSSHPANMHLGMSLLETDCGGFM 122
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 71 GFPENRMGDVYKCPV-DLSTATCEKLNQTSIPNVTMTKNSLGLILTRNMTGCGFL 129
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 123 ACAPLWSRACGSSVFSGGICARVDASFPQCSLAPTAQRCPYMDVIVLDGNSIYPS 182
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 130 TCGPLWAQCCGNYTITGVCSDISPDFQLSASFATQPCPSLDIVVVCDESNIYPWD 189
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 183 EVQTFLLRLVGLKFLDPEIQVGLVQYGESVPVHWSLGDFTKEEVRRAAKNLSRREGRE 242
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 190 AVKNFLEKFGVQLDIPGKTQVGLIQYANPRVFNLTNTYKTEEMIVATSQTSQYGGDL 249
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 243 TKTAQAIAWACTEGFSQSHGGRPEARLLVVVTDGESHGDELPALAKACBAGRVTYGI 302
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 250 TMTFGAIQYARKYAYSAAGSGRRSATKVMVVVTDGESHGSMKAKAVIDQCNDHNLRFGI 309
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 303 AVLGHYLRQRDPSPFLREITIASDPERFFNVTDDEAALTDIVDALGDRIFGLESHA 362
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 310 AVLGYLARNALDXTNLKAIKAIASIPTEYFFNVSDAALLKAGTLEQIFSIETG-V 368
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 363 ENESFGLMSQIGFST--HRLKDGILFGMVADWGGSVLW-LEGHRLPPRMALEDE 419
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 369 QGDNFQWMSQVGSFADYSQNDILMLGAVAGFWSGTIVOKTSHGLIFP-----KQA 423
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 420 PPPALQ--NHAAYLYGVSSMLRGRRLPLSGAPRFRHRKGVIAFOLKKDGVAVRVAQSL 477
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 424 FDQILQDRNHSYLYGSVAA-ISTGESTHFVAGAPRANYTQOIVLYSWENGNITVIAH 482
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 478 QGEQIGSYFGSELCPDLTDRTGTDVLLVAAFMFLGPONKETGRVYVY-----LVGQSL 532
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 483 RGDQIGSYFGSVLCSDVDKDTITDVLVAGAPMTMSDLKKEEGRVYLFITKKGILGHQF 542
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 533 LTLQGLTLOPEPPQDARFGFAGALPDNLNODGFADVAVGAPLEDHOGALYLYHGTQSGVR 592
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 543 --LEG---PEGIENTRFGSAALSDINMGDFNVVIGSPLENQNSGAVIYINGHQGTIR 597
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 593 PHPAQRIAAA--SMPHALSIFGRSDVGRDLDDLDVAVGAQGAAILLSRRPIVHLTP 650
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 598 TKYSQKILSGDGAFRSHLOVYGRSLDGYGLDNGSDITDVSIGAFQVQVQLWSQSIADVAI 657
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 651 SLEVTPOAISVVQDRCRRRQGEAVCLTALCFQVTSRTPPGRWHDQFYMRFTASLDIEWTAG 710
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 658 EASPTPEKITLVNNAQ-----IILKCFSAKFR--PTKQNNQVAIVNITLDA----- 704
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 711 ARAAPDGGORLSPRRL-----RLSVGNV-----TCEQLHFHVLDTSDYLRPVALTVT 758
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 705 -----DGFSSRVTSGRLPKENNERCLQKNMVVNAQSCPEHIYIQEPSDVVNSLDURVD 759
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 759 FALDNTTKPG--PVLNIGSPTSIOKLVPFSKDCPDNECVTLDLVLQNMNDRGRKAPFV 816
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 760 ISLEN---PGTSPALAEVSETAKVFSIPFKDCGDEGLCISDLVLDVR-QIPAAQEQPFI 815
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 817 VRGGRKVLVSTTLNKRKENAYNTSLIIFSRNLHLASLTQORESPI---KVEC-APPSA 872
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 816 VSNQKRLTFSTVLKNKESAYNTGIIVDFSENLPFFASFS---LPVDGTETVTCQVAASQ 871
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 873 HARLCSVCHPVFQTCAKVTLELEFEFSCSLLSQVFGKLTASSDSLENGTLQENTACTS 932
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
US-09-949-016-10747
; Sequence 10747, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10747
; LENGTH: 1195
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10747

Query Match      28.8%; Score 1756; DB 4; Length 1195;
Best Local Similarity 34.3%; Pred. No. 3e-160;
Matches 418; Conservative 219; Mismatches 455; Indels 126; Gaps 31;

Qy 11 LPLVFLTGL-----CSPFNLDHHPRLFPGPPEAFSGYVLQHVGGQRMVLVGPWD 63
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 25 LPLLLVLSOGILNCLLAYNVGLPEAKIFSGPSEQFGYAVQVQFINPKGNWLLVGPWS 84
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 64 GPSGDRGDVYRCPVGGAHNAPCAKHL-GDYQGNSSHPANMHLGMSLLETDCGGFM 122
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 85 GFPENRMGDVYKCPV-DLSTATCEKLNQTSIPNVTMTKNSLGLILTRNMTGCGFL 143
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 123 ACAPLWSRACGSSVFSGGICARVDASFPQCSLAPTAQRCPYMDVIVLDGNSIYPS 182
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 144 TCGPLWAQCCGNYTITGVCSDISPDFQLSASFATQPCPSLDIVVVCDESNIYPWD 203
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 183 EVQTFLLRLVGLKFLDPEIQVGLVQYGESVPVHWSLGDFTKEEVRRAAKNLSRREGRE 242
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 204 AVKNFLEKFGVQLDIPGKTQVGLIQYANPRVFNLTNTYKTEEMIVATSQTSQYGGDL 263
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 243 TKTAQAIAWACTEGFSQSHGGRPEARLLVVVTDGESHGDELPALAKACBAGRVTYGI 302
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 264 TMTFGAIQYARKYAYSAAGSGRRSATKVMVVVTDGESHGSMKAKAVIDQCNDHNLRFGI 323
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 303 AVLGHYLRQRDPSPFLREITIASDPERFFNVTDDEAALTDIVDALGDRIFGLESHA 362
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; 
```

```
Db 324 AVGLYLRNALDTKNLIKETAISIPTRYFFNVSDAAALKEKAGTLGQIFSIET-V 382
Qy 363 ENBSFGLMSQIGFST--HRLKDGILFGVAVDMGGSVLM-LEGHRLPPPPMALEDE 419
Db 383 QGGDNFQWMSQVGSADYSQNDILMLGAVGFWGSGTIVQKTSHGHLIFP-----KQA 437
Qy 420 FPPALQ--NHAAYLYGVSVMMLRGRRLLFLSGAPRFRHRGKVIQFOLKDKDGVAVRVAQSL 477
Db 438 FDQLQDRNHSSYLYGSVAA-ISTGBSTHFVAGAPRANYTGOVLVSVNENGNIIVLQAH 496
Qy 478 QGBOIGSYFGSELCPDLTDRDGTDLVLLVAPMFLGPONKETRGRVYVY-----LVGQOQL 532
Db 497 RGDQIGSYFGSVLCSVDVDDTITDVLVAGAPMTMSDLKKEBGRVYLFTIKEGILGQHOF 556
Qy 533 LTQGLTLQPPPPQDARFGFAMGALPDINQDGFADVAVGAPLEBQHQCALVLYHCTQSGVR 592
Db 557 --LEG-----PGIENTRFGSAIALUSDINMGDFNDVIIVGSPLENGSAGVIYNGHQTIR 611
Qy 593 PHPAQRIAAA--SMPHALSFGRSVDGRLDLGDGDLVDVAVGAQAAILLSSRPVHLTP 650
Db 612 TKYSQKILSGDGAFRSHLQFGRSLDGYDGLNGDSITDVISGAFQGVVQLWSQSIADVAI 671
Qy 651 SLEVTPOAISVVQDRCRRRQGEAVCLTAALCFQVTSRTPGRWDHQFVNRFTASLDEWTAG 710
Db 672 EASFTPEKITLVNKAQ-----IILKLCFSAPFR-PTKQNNQVAIVYNNITLDA---- 718
Qy 711 ARAAFDGGGOLSPRLR-----RLSVGNV-----TCEQLHFHVLDTSDYLRPVALTVT 758
Db 719 -----DGFSSRVTSRGLFKENNERCLQNMVYVNAQSCPBEHIIYIQBPDVNVSLDURVD 773
Qy 759 FALDNTTKPG--PVLNBSGSPTSIQKLVFPGKDCPDNECVTDLVLVQNMIDIRSRKAPFV 816
Db 774 ISLEN-----PGTSPALEAYSETAKVFSIPFHKDCGEDGLCISDLVDYR-QIPAAQEQPFI 829
Qy 817 VRGRRKVLVSTTLNENKENAYNTLSIIIFSRNLHLASLTPORESPI---KVEC-AAPSA 872
Db 830 VSNQKRLTFSVTLKKNRESAYNTGIIVDFSENLFASFS----LPVDGTETVCQVAASQ 885
Qy 873 HARLCSVGHVPVPGAKVTLLEPFSCSLLSQVFGKLTASSDSLERNGTLQENTAQTS 932
Db 886 KSAACDVGYPALKEEQVFTFINDFNLQNLQNASISFQALSSESQENKA--DNLVNLK 943
Qy 933 AYIQEYELLHPSSESTLHREVEHPYGTLP-----VGPGEFKTTLRVQLGCVVWSGLI 986
Db 944 IPLYDAEIHLTRSTNFINFEISSDGNVPSIVHSEFEDVGPKEFISLKV-TTGSVPVSMAT 1002
Qy 987 ISALLPAVAGGNVFLSLSQVITNNASCIQNLTEPP-----GPPVHPPELOHTNRL 1038
Db 1003 VIIHIPQVTKENPLMYLTGVQTDKAGDISCNADINPLKIGQTSSTSSVSFKSENFRHTKEL 1062
Qy 1039 NGSNTQCOVVRCHLQOLAKGTEVSUGLLRLVHNEFFRRAKFKSLTVVSTPELGTTEGSVL 1098
Db 1063 NCRTASCNTVCLWKDVMHKGSEYFVNTTIRWNGTFASSTFTQVLTAAABINTPEIY 1122
Qy 1099 QLTEASRWSESLLEVQOTRPLISLWI-----LIGSVLGLLILLALLVPCWLW 1145
Db 1123 -----VIEDNTVTIPLMIMKPEKAEVPTGVIGSIIGIAGILLALLVAILW 1168
Qy 1146 KLGPFAHK-----KIPBE 1158
Db 1169 KLGFPRKRYEKWTKNPDE 1186
```

## RESULT 7

US-09-532-310B-6

Sequence 6, Application US/09532310B

Patent No. 6596276

GENERAL INFORMATION:

APPLICANT: Senger, Donald R

Detmar, Michael

Claffey, Kevin P

TITLE OF INVENTION: Method for inhibiting tumor

```
angiogenesis in a living subject
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: David Prashker, Esq.
STREET: P.O. Box 5387
CITY: Magnolia
STATE: Massachusetts
COUNTRY: USA
ZIP: 01930
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage
COMPUTER: Dell PC
OPERATING SYSTEM: MS DOS
SOFTWARE: Microsoft Word version 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/532,310B
FILING DATE: 22-Mar-2000
CLASSIFICATION: Unknown
ATTORNEY/AGENT INFORMATION:
NAME: David Prashker, Esq.
REGISTRATION NUMBER: 29,693
REFERENCE/DOCKET NUMBER: BIS-036
TELECOMMUNICATION INFORMATION:
TELEPHONE: (978) 525-3794
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1183 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-532-310B-6
```

Query Match 28.3%; Score 1728; DB 4; Length 1183;

Best Local Similarity 33.7%; Pred. No. 1.5e-157;

Matches 412; Conservative 229; Mismatches 459; Indels 124; Gaps 32;

Qy 11 LPLVFLTGL-----CSPENLDEHHPRLPPGPEAEFGYSVLQHVGGGQRMVLVGPWD 63

Db 11 LPLVLLVLSQGLINCLAYNLVGLPEAKIFSGSPSSQFGYAVQOQINPKGNMLLVGSPWS 70

Qy 64 GPSDGRRGDYVRCPVCGAHNAPCAKQHL-QDYQLGNSHPAVNMHLGMSLLETDDGGGPM 122

Db 71 GPFENRMGDVYKCPV-DLSTATCEKLNLTQSTSPNVTEMKTNMSLGLLTRNMGTCGFL 129

Qy 123 ACAPLWSRACSSVFSGGICARVDASFQPGSLAPTAQRCPTVMDVVIVLDGNSIYVWS 182

Db 130 TCGPLMAQQCGNYTTGVCSDISPDFXKSASFSPATXPCPSLIDVVVVCDESNIYVWD 189

Qy 183 EVOTFLRLVLGKLFIDPEIQVGLVOYGESPVHWSLGDFTKBEVVRAAKNLSRRRGRE 242

Db 190 AVKNLEKFFVQGLDIGTKTQVGLIYANNPRVFNLTYYTKTEEMIVATSTQSYQGD 249

Qy 243 TKTAQAIMVACTBFGSQSHGRPEARLLVVYVTDGESHDEELPAALKACEAGRVTYGI 302

Db 250 TMTFAGIYQARKYVSAASGRRSATKVMVVTVDGESHGSMKKAVIDQCNDHNLRFGI 309

Qy 303 AVLGHYLRORDPS-SFLREIRIYASDPDRPFNVYDDEAALTDIVDALGDRIFGLESH 361

Db 310 AVLYLNRNALDTKNMLIKEIKATASIPTRYFFNVSDAAALKEKAGTLGQIFSIET- 368

Qy 362 AENESSFGLMSQIGFST--HRLKDGILFGVAVDMGGSVLM-LEGHRLFFPRM--AL 416

Db 369 VGGDNFXMWSQVGSADYSQNDILMLGAVGAFGWSGTIVQKTSHGHLIFPKQAQDOI 428

Qy 417 EDEFPFPALQNHAAAYLYGYSVSSMLLRGRRLLFLSGAPRFRHRGKVIQFOLKDKDGVAVRVAQS 476

Db 429 QD-----RNHSSYLYGSVAA-ISTGBSTHFVAGAPRANYTGOVLVSVNENGNIIVQIA 481

Qy 477 LQGEQIGSYFGSELCPDLTDRDGTDLVLLVAPMFLGPONKETRGRVYVY-----LVGQOQL 531

Db 482 HRGDQIGSYFGSVLCSVDVDDKDTITDVLVAGAPMTMSDLKKEBGRVYVLTIKKILGQHQ 541

QY 532 LLTLQGTLOPEPPQDARFGFANGALPDLNODGFADVAAPLEDHCHQALYLYHGTSQGV 591  
Db 542 F--LEG---PEGIENTFGSAIALSINMDGDFNDVIVGSPLENQNSGAVIYNGHGTI 596  
QY 592 RPHPAQRIAAA--SMPHALSIFYGRSVDGRDLDDGDLVDVAVAGAGAAILLSSRPVHLT 649  
Db 597 RTKYSQKILSGDGAFAHSLQYFGSLDGYDGLNGDSITDVISGAFGVQVQLWSQSIADVA 656  
QY 650 PSELTVOAISVQDRCRRQGEAVCIATAALCFQVTSRTPGRWDPHQPMRFTASLDWETA 709  
Db 657 IEASFTPEKITLVNNAQ-----IILKLCFSAKFR-PTKONQVAIVNITLDA--- 704  
QY 710 GARAADGSGORLSPRL-----RLSVGNV-----TCEQLHFHVLDTSLRPV-ALT 756  
Db 705 -----DGFSSRVTSRGLFKENNERCLOKXNVVNOAQSCPEHIYIQEPSDVNSLDDL 758  
QY 757 VTFALDNTTPG--PVLNKGSPGPTGIQLVPPSCDGPNECVTDVLQVNNMDIRGSRKAP 814  
Db 759 VDISEN---PGTSPALEAYSETAKVPSIPPHKDCGEGKISDLVQ-DVRIAPAAQEQP 814  
QY 815 FVVRGGRKVLVSTLENKRNAYNTSLISIFSRNLHLASLTPORESPI---KVEC-AAP 870  
Db 815 FIVSNONKRLTFSVTLKNKRESAYNTGIVVDFSENLFASFS---LPVDGTEVTCQVAA 870  
QY 871 SAHARLCVGHVPQTGAKVTFLFEPFSCSSLLSQVFGKLTASSDSLERNGTLOENTAQ 930  
Db 871 SQKSVACDVGYPALKRQEQVTFITFDFNLQNLQNASLSFQALSSESQENKA---DNLVN 928  
QY 931 TSAVIOVEPHLLFSESSTLHRYEHPVGTLP-----VGPPEPKTTLRVQNGLCYVVS 984  
Db 929 LKIFLLDAEHLTRSTNINYEISSDGNVPSIVHSFEDVGPKFIFSLKV--TTSQVPSVM 987  
QY 985 LIISALLPAVAGNGYFLSLSQVITNNASCIVQNLTEPP-----GPPVHPBELOHTN 1036  
Db 988 ATVIIHIPQVTKENKPLMYLGVQTDKAGDCCNADINPLKIGOTSSSVSFKSENFRHTK 1047  
QY 1037 RLNGSNTQCVVRCHLQLAGTEVSGLLRLVINEFFRAKFKSLTVSVTFELGTBEGS 1096  
Db 1048 ELNCRATSCSNVTCLWLDVHMKGIFYFNVVTRIWNGTFASTFTQVLTAAAEINTYNPE 1107  
QY 1097 VLQUTEASRWSSELEVVOTRPIILSLMI-----LIGSVLGGILLALLVFC 1143  
Db 1108 IY-----VIENTVTIPLMIMKPDKEAEVPTGVIIGSIAGILLALLVAI 1153  
QY 1144 LWKLGFFAHHKIPBEEREKLEQ 1167  
Db 1154 LWKLGFFKRYEKWKTKNPDEIDE 1177

## RESULT 8

US-08-173-497-2  
; Sequence 2, Application US/08173497  
; Patent No. 5437958  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van Der Vieren, Monica  
; TITLE OF INVENTION: No. 5437958el Human 2 Integrin Alpha  
; TITLE OF INVENTION: Subunit  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 S. Wacker Drive, 6300 Sears Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/173,497  
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: No. 5437958and, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 27866/31363

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1161 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-173-497-2

Query Match 18.5%; Score 1131.5; DB 1; Length 1161;

Best Local Similarity 29.4%; Pred. No. 1e-99;

Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

QY 23 FNLDEHHPRLPFGPPEAEFGVSLQHYGGORWMLVAPWDPGSDRRGVYRCPVCGAH 82

Db 17 FNLDVVEPTIQ-EDAGGFGQSVVQF--GGSRLVVGAPLEVVVAANTGRLYDC---AAA 69

QY 83 NAFCAKGHLDGYQLGNSHPAVNMHGLMSLLETDGCGFMACAPLWMSRACGSSVFSGIC 142

Db 70 TGMCOPIPL-----HIRPEAVNMSLGLTLAASNGSRLLACGPTLHRVCGENSYSKGC 123

QY 143 ARVDASFPQGSALAPTAQRCPT-TYMDVVIVLDGNSI--YPMSEVQTFRLRLVKGLFIDP 199

Db 124 LLLGSRWEIITQVTPDAPTECPHOEMDIVFLIDGSGSIDQNDFNQMGKFVQAVMQQ--FEG 181

QY 200 EQIQVGLVQVQGESVPHWMSLGDRTKEEVVRAAKNLSRREGRETKTAQAAMVACTEFSQ 259

Db 182 TDTLFAIMQVSNLKIHTFTQRTSPSQSLVDPIVQLKGL-TFTATGILTVVTVQLFHH 240

QY 260 SHGRPEAARLLVVVTDGESH-DGEELPAALAKACEAGRVTYRGIYAVLGHVLRORDPSSF 318

Db 241 KNGARKSAKILVITDQKQKPLEYSDVIPAQKAGIIRYAIQV-GHAF---QGPTA- 295

QY 319 LREIRITASDPDRFFNFVTVDEAALTDIADALGRIFGLEGSHAENESSFLEMSQIGFS 378

Db 296 ROELNTISSAPPQDHVFKVDNFAALGSIQKLOEKIYAVEGTQSRASSSFQHEMSQEGFS 355

QY 379 THRLKDGILFGMGAYDMGCVLWLEGGHRLFPPMALEDEFFPALONHA-----AY 430

Db 356 TALTMGDLFLGAVGFSFSW-----SGGAFLYPPNMS-----PTFINMQENVDMRDSY 402

QY 431 LGYSVSMILLRGRRFLSGAPRFRHRGKVIAP-QLKKDGAVRVAQSLQGEIQIGSYFGSE 489

Db 403 LGYSTELALWKGVQNLVL-CAPRYQHTGKAVITQVSRQW--RKAEVTTGTQIGSYFGAS 459

QY 490 LCPDLTDRDGTTLVLLVAAPMFLGPQNKETGRVYVYLV--GQOSLLTLQGTLOPEPPQD- 546

Db 460 LCSVDVDSGSDTLILIGAPHYV--EOTRGQVSVCPPLPRGQVQVQWCDVALRGEQHPW 517

QY 547 ARGFAMGALPDLNODGFADVAAPLEDHCHQALYLYHG-TQGVTPHAPQRIATAASMP 605

Db 518 GRFGAALTVLGDVNEKLDIIVAI GAGEQENRGAVLFGHASESGISPSHSQRASSQLS 577

QY 606 HALSYFGRSVDGRDLDDGDLVDVAVAGAGAAILLSSRPVHLTPSLEVPQAISVVQRD 665

Db 578 PRLOYFGQALSGGODLTQDGLMDLAVGARGQVLLRLSLPVLKVCVAMRFPSEPVAKAYR 637

QY 666 C-----RRRQGEAVCLTAALCFQVTSRTPGRWDPHQVYMRFTASLDWETAGAPAAFDGS 718

Db 638 CWEEKPSALEAGDATVCLTIQ-----KSLDQLDGIQSSVRFDLALDPLGRILTSRAIFNET 692

QY 719 GQRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRLPVALTVTTFALDNTTKPG-----PVLN 772

Db 693 KNPTLTRKTLGLG-IHCETLKLDPCEVDVWSPILHLNLSLVRBPFPSPQNLRPVLA 751  
Qy 773 EGSPTSIOKLVPFSGDCGPNCEVTDLVLQNMDIRSKAPFVVRGRKVLVSTTLEN 832  
Db 752 VGSQDLFTASLPFFKNGQDGLCEGD--LGVTLFSGLQT---LTVGSSLELNVIVTWN 806  
Qy 833 RKENAYNTSLIIPSRNLHLASLTPQRESP-----IKVEC-AAPSAHARL-----CSVGHPV 883  
Db 807 AGEDSYGVTVSLYYPAGLSHRRVSGAQKQPHQSALRLACETVPTDEGLRASSRCSVNHPI 866  
Qy 884 FOTCAKVTFLLEFPFSCSSLLSQVFGKLTASSDRLRNGTLQENTATQSAIYQEPHLF 943  
Db 867 FHEGNGTFTVTFDVSYSKATLGDRLM-LMRASSENKASSKATFQLELFPKYAVTMI 925  
Qy 944 SSESTLHYEVHPYGTLPVGPGRFETTLRVNLCVGVVSGLLIISALLPAVAHGNVFLS 1003  
Db 926 SRQESTKY--FNPATSDKKMEAEHRYRVNLSQRDLA-ISINFVPLVLLNGAVAVMDV 982  
Qy 1004 LSQVITNNAACIVQNLTEPPGPPVHPBELQHTNR---LNGSNTQCQVVRCHLQAKGTE 1060  
Db 983 VMEAPSQSLPCVSR-----KPPQHSDFLTQISRSPMLDCSIADCLQFRCDVPSFSVQEE 1037  
Qy 1061 VSVGLLRLVNEFPFRRAKFKSLTVSTFELGTEGVSQVLTQTEASRWSESLEL-VVQTRPI 1119  
Db 1038 LDFTLKNLSFGWVRETLOKKVLVSVAEITFDTSVTSQLPQGFAPFRAQMEMVLEBDEV 1097  
Qy 1120 LISLWILIGSVLGLLALLLALVCLWKLGPFAHKKIPEREKREE 1163  
Db 1098 YNAIPIWGSSVGLLALLLALITATLYKLGFKRHYKMLEDKPED 1142

RESULT 9  
US-08-286-889-2  
; Sequence 2, Application US/08286889  
; Patent No. 5470953  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Mich  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: P38,659  
; REFERENCE/DOCKET NUMBER: 27866/32168  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1161 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-286-889-2  
Query Match 18.5%; Score 1131.5; DB 1; Length 1161;  
Best Local Similarity 29.4%; Pred. No. 1e-99;  
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;  
Qy 23 FNLDHHPRLFPQPPBAEFYSVLQHVGGQRMMLYGAPMDGSPGDRRGDVRCVPVGAH 82  
Db 17 FNLDVEEPTIFQ-EDAGGFGQSVVQF--GGSRLVVGAPLEVVAANQTGLYDC---AAA 69  
Qy 83 NAPCAKXGHLGDYOLGNSSHPAVNMHGLMSLLETDDGGGFMACAPLMSRACGSSVPSGIC 142  
Db 70 TGMCOPIPL-----HIRPEAVNMSLGLTLAASNGSRLLACGPTLHRVCGENSYSKGSC 123  
Qy 143 ARVDASFPQGSAPLAPTAQRCP--TYMDVVLVLDGNSI--YPMSEVQTELLRLVGLFIDP 199  
Db 124 LLGSRWEIITQVPDATPECPHQMEDIVLIDGSGSIDQNDFNQMGKGFQAVNMQ--FEG 181  
Qy 200 EQIQVGLVQYGESPVHWSLGDPTKEEVVRAAKNLSRREGRETKTAQIMVACTEGFSQ 259  
Db 182 TDTLPALMQYSNLLKIHTFTPTQRTSPSQSLVDPVQLKGL-TPTATGILLTVVTQLFHH 240  
Qy 260 SHGGRPEAAKLLVVTDGSSH--DGBELPAALKACEAGRVTRYGIAGVLGHVLRQRDPSSF 318  
Db 241 KNGARKSAKKILVITDQGYKDPLEYSVDVPOAEKAGIIRYAIGV--GHAF---QGPTA- 295  
Qy 319 LRIRTIASDPDERPFEPFNVTDEAALTDIVDALGDRIFGLGSGHAENESSFGLMSQIGFS 378  
Db 296 RQELNTISSAPPQDHVFKVDNFAALGSIQKQLEKIYAVEGTOSRASSSPQHMSQEGFS 355  
Qy 379 THRLKDGILFGVMGAYDWGGSVLMLEGGHRLFPFRMALEDEFPFPAQLNHA-----AY 430  
Db 356 TALTMGDLFLGAVGSPFSW-----SGGAFLYPNMS-----PTFNNMSQENVMDRDSY 402  
Qy 431 LGYSVSMLLRGRRRLFLSGAPFRHRGKVIAF-QLKXGAVRVAOSLQEQIGSVFGSE 489  
Db 403 LGYSTELALWKGVQNLV--GAPRYQHTKRAVITQVSROW--RKAENVGTQIGSYFGAS 459  
Qy 490 LCPDTRDGTDTDLVAAAPMFLGPQNKETGRVYVVLV--GQOSLLTLQGTLOPEPQD- 546  
Db 460 LCSVDVDSGSDTLILIGAPHYY--BOTRGQVSVCLPRGQVQWQCDAVLGEGQHPW 517  
Qy 547 AREGFAMGALPDNLQDGFADVAVGAPLEDHQCALYLYHG--TQSGVRPHPAQIAAASMP 605  
Db 518 GREGAALTVLGDVNEKLDLVAIGAPEGQENRGAVVLPHGASGSGISPHSQRIASSQLS 577  
Qy 606 HALSYFGRSVDRDLDDGDDLDVDAVGAQGAAILLSRPIVHLTPSLEVTTPQAISSVQRD 665  
Db 578 PRLQYFQALSGGQDLTQDGLMDLAVGARGQVLLRLSLPVLKVGAMRFSPEVAKAVYR 637  
Qy 666 C-----RREGQEAVALTAALCFQVTSRTPGRWHDQFYMRFTASLDENTAGARAAPDGS 718  
Db 638 CWEEKPSALUEAGDATVCLTIQ-----KSLDQGLDIQSSVRFDLALDPGLTSLRAIPNET 692  
Qy 719 GQRLSPRRRLKLSVGNVTCEQLHFPVLD--TSDYLRLPVALTFTFALDNTTKFG-----PVLN 772  
Db 693 KNPTLTRKTLGLG-IHCETLKLDPCEVDVWSPILHLNLSLVRBPFPSPQNLRPVLA 751  
Qy 773 EGSPTSIOKLVPFSGDCGPNCEVTDLVLQNMDIRSKAPFVVRGRKVLVSTTLEN 832  
Db 752 VGSQDLFTASLPFFKNGQDGLCEGD--LGVTLFSGLQT---LTVGSSLELNVIVTWN 806  
Qy 833 RKENAYNTSLIIPSRNLHLASLTPQRESP-----IKVEC-AAPSAHARL-----CSVGHPV 883  
Db 807 AGEDSYGVTVSLYYPAGLSHRRVSGAQKQPHQSALRLACETVPTDEGLRASSRCSVNHPI 866  
Qy 884 FOTCAKVTFLLEFPFSCSSLLSQVFGKLTASSDRLRNGTLQENTATQSAIYQEPHLF 943  
Db 867 FHEGNGTFTVTFDVSYSKATLGDRLM-LMRASSENKASSKATFQLELFPKYAVTMI 925  
Qy 944 SSESTLHYEVHPYGTLPVGPGRFETTLRVNLCVGVVSGLLIISALLPAVAHGNVFLS 1003  
Db 926 SRQESTKY--FNPATSDKKMEAEHRYRVNLSQRDLA-ISINFVPLVLLNGAVAVMDV 982



Db 1098 YNAIPIIMGSSVGGALLLALITATLYKLGFFKHYKEMLEDPED 1142

RESULT 11

US-08-362-652-2

; Sequence 2, Application US/08362652

; Patent No. 5766850

GENERAL INFORMATION:

APPLICANT: Gallatin, W. Michael

APPLICANT: Van der Vieren, Monica

TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit

NUMBER OF SEQUENCES: 93

CORRESPONDENCE ADDRESS:

ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 233 South Wacker Drive, 6300 Sear Tower

CITY: Chicago

STATE: Illinois

COUNTRY: United States

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/362,652

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/173,497

FILING DATE: 23-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/286,889

FILING DATE: 5-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.

REGISTRATION NUMBER: 38,659

REFERENCE/DOCKET NUMBER: 27866/32391

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1161 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-362-652-2

Query Match 18.5%; Score 1131.5; DB 1; Length 1161;

Best Local Similarity 29.4%; Pred. No. 1e-99;

Matches 346; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

Qy 23 FNLDHPRLPFGPPEAEFGYSLVQHGQGMWLVGAPMDGPGDRRGDYRCVGGAH 82

Db 17 FNLDVEPTIFQ-EDAGFGQSVVQF--GGSR-LVVGAPLEVAANQTGRLYDC--AAA 69

Qy 83 NAPCAKHLGDYQLGNSSHAPVANNHGLMSLLETGDCGFMACAPLWSRACSSVSSGIC 142

Db 70 TGMCOPIPL-----HIRPEAVNNSLGLTLAASTNGSRLLACGPTLHRVCGENSXKGC 123

Qy 143 ARVDASFPQGSGLAPTAQRCP--TYMDVVIVLDGNSI--YPMSEVQTFRLRLVCKLFDIP 199

Db 124 LLLGSRWEIITVPDAPTECPHQEMDIVFLIDGSGSIDQDNFNQMKGFVQVMQ--FEG 181

Qy 200 EQIQVLVQYGESPVHMSLGDGFRTEKEVWRAAKNLSRREGRETKTQAQIMVACTEGFSQ 259

Db 182 TDTLFPALMWSNLLKIHFTFTQFRTPSQSGLVDPPIVQLKGL-TFTATGILTVVTQLFHH 240

Qy 260 SHGRPEARLLVVVTDGESH-DEEELPAALKACEAGRVTRYGTAVLGHYLRORDPSSP 318

Db 241 KNGARKGAKKILIVITDGQKYKDPLEYSDDVIPQAEKAGIIRYAIGV-GHAF--QGPTA- 295

Qy 319 LRBIRTIASDPDRFFNFVTDEAALTDIVDALGDRIFGLEGSHAENESSFGLMSQIGFS 378

Db 296 RQELNTISSAPPQDHPVKVDNFAALGSIQKQLOEKIYAVEGTQSRASSSFQHEMSQEGFS 355

Qy 379 THRLKDGILFGMWGAYDWGGSVLMWEGHRLFPFPRMALEDEFPFPAQONHA-----AY 430

Db 356 TALTMGDLFLGAVGSPW-----SGGAFLYPPNMS-----PTFINNSQENVMDRDSY 402

Qy 431 LGYSVSSMLLRGGRRFLSGAPRFRHKGVIAP-QLKKGAVRVAQSLQGEQIGSYSGSE 489

Db 403 LGYSTELALMWGVONLVL-GAPRYOHTGXAVFTQVSQW--RKKAETVGTQIGSYFGAS 459

Qy 490 LCPDLTDRGTTDVLVLAAPMFLGPQNKETGRVYVYL--GQOSSLTLQGLTQLEPPQD- 546

Db 460 LCSVDVDSGSDTLILIGAPHYY--BQTRGGQSVSCPLPRGQVQWQCDVLEGEQHPW 517

Qy 547 ARFGFAMGALPDINQDGPADVAVGAPLEBHQALYLYHG-TOSGYRPHPAQRIAAASMP 605

Db 518 GRFGAALTVLGDVNEKLDLDAIGAPEGQENRGAVYLFHGASBSGISPSHSQRIASSQLS 577

Qy 606 HALSYFGRSVDGRLDLDGDDLDVDAVGAQGAAILLSRPPIVHLTPSLVTPQAISVVQD 665

Db 578 PRLOYFGQALSGGQDLTQDGLMDLAVGARGQVILLRSLPVLKVGAVMRFPSPVEKAVYR 637

Qy 666 C-----RRRQGEAVCLTAALCFQVTSRTPGRWHDHQFYMRFTASLDEMTAGARAAFDGS 718

Db 638 CWEEKPSALEAGDATVCLTIQ-----KSSLDQGLDIQSSVRFDLALDPGLTSAIFNET 692

Qy 719 GQRLSPRLRLSVGNVTCBQLPHVLD-TSDYLRPVALTVPALDNTTKPG-----PVLN 772

Db 693 KNPTLRRKTLGLG-THCETLKLLPDCVEDVVSPIILHLNLSLREPIPSQNLRLPVLA 751

Qy 773 EGSPTSIOQLVPFSKDCGPDNECVTLVLQVNMDIRSKAPFVVGRRKVLVSTLEN 832

Db 752 VGSQDLFTASLPFEKNCQDGLCEGD--LGVTLSPSGLQF---LTVGSSLELNVIVTVN 806

Qy 833 RKENAYNTSLIIFSRNLHSLTPORESF-----IKVEC-AAPSAHARL-----CSVGHVP 883

Db 807 AGEDSGTVTVSLYYPAGLSHRRVSGAQKQPHQSALRLACETVTEDEGLSSRCVSNHPI 866

Qy 884 FOTGAKVTLLFEFSCSSLLSQVFGKLTASSDSLRNGTFLQNTAQTSAIYIEPHLLF 943

Db 867 FHEGSGTFTVTFDVSFKATLGDRM-LMRASASSENKASSKATFQLELPVKYAVVTMI 925

Qy 944 SSESTLHRYEHPYGTLPVGPGEFTLVRQNLGCVVVSGLLIISALLPAVAGHGVFLS 1003

Db 926 SROEESTKY--FNFATSDKKMKAEHRYRVNLSQORDLA-ISINFVPLVLLNGVAVMDV 982

Qy 1004 LSQVITNNASCIQNLTEPPGPPVHPEELQHTNR---LNGSNTQCVVRCHLQOLAKGTE 1060

Db 983 VMEAPSQSLPCVSR-----KPPQHSDFLTQSRPMLDCSIADCLQFRCDDVSPSVQEE 1037

Qy 1061 VSVGLRLVHNEFFRRAKFKSLTVVSTFELGTBEGSVLQLTASRWSESLLE-VVQTRPI 1119

Db 1038 LDFTLKNLSFGWVRRETQKKLVWSVAEITFTDTSVYSQLPGQEAQFMAQMEMVLEDEV 1097

Qy 1120 LISMLILIGSVLGGALLLALLVFLCKLGPFAHKKIPEEKKEE 1163

Db 1098 YNAIPIIMGSSVGGALLLALITATLYKLGFFKHYKEMLEDPED 1142

RESULT 12

US-08-605-672-2

; Sequence 2, Application US/08605672

; Patent No. 5817515

GENERAL INFORMATION:

APPLICANT: Gallatin, W. Michael

APPLICANT: Van der Vieren, Monica

TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit

NUMBER OF SEQUENCES: 103

CORRESPONDENCE ADDRESS:

ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun

```

, STREET: 233 South Wacker Drive, 6300 Sear Tower
, CITY: Chicago
, STATE: Illinois
, COUNTRY: United States
, ZIP: 60606-6402
, COMPUTER READABLE FORM:
, MEDIUM TYPE: Floppy disk
, COMPUTER: IBM PC compatible
, OPERATING SYSTEM: PC-DOS/MS-DOS
, SOFTWARE: Patent In Release #1.0, Version #1.25
, CURRENT APPLICATION DATA:
, APPLICATION NUMBER: US/08/605,672
, FILING DATE:
, CLASSIFICATION: 530
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 08/173,497
, FILING DATE: 23-DEC-1993
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 08/286,889
, FILING DATE: 5-AUG-1994
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US 08/362,652
, FILING DATE: 21-DEC-1994
, ATTORNEY/AGENT INFORMATION:
, NAME: Williams Jr., Joseph A.
, REGISTRATION NUMBER: 38,659
, REFERENCE/DOCKET NUMBER: 27866/32684
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: 312-474-6300
, TELEFAX: 312-474-0448
, TELEX: 25-3856
, INFORMATION FOR SEQ ID NO: 2:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 1161 amino acids
, TYPE: amino acid
, TOPOLOGY: linear
, MOLECULE TYPE: protein
, US-08-605-672-2

```

Query Match	18.5%	Score	1131.5	DB 2	Length	1161			
Best Local Similarity	29.4%	Pred. No. le-99							
Matches	348	Conservative	215	Mismatches	519	Indels	103	Gaps	40

  

Qy	23	FNLDEHHPRLPFGPPEAEFGVSVLQHVGGORWMLVGAPWDGPGSDRRGDVYRCPVCGAH	82
Db	17	FNLDSVEPTIQ--EDAGFGOSVQF--GGR-LVVGAPLEVAANOQTRYDC---AAA	69
Qy	83	NAPCAKHLGDDYQLGNSHSPAVNMHLGMSLLETDDGGFMACAPLWSRACGSSVFSSGIC	14
Db	70	TGMCOPIPL-----HIRPEAVNMSLGLTAASTNGSRLLACGPTLHRVCGENSYSKGC	123
Qy	143	ARVDASTQPGSLAPTQRCP-TYMDVVVLVDGNSI--YPMSEVQTFRLRLVGLKLFIDP	199
Db	124	LLLSGRWEIIQTVDATPECPHQEMDIVFLIDGSGSDQDNDFNMKGFGVQAVMQQ--FEG	181
Qy	200	EQIQVGLVQVGESPVHWSLGDPTKEEVRAAKNLRRREGREKTAQATMVACTEGFSQ	259
Db	182	TDTLFALMQYNLKIHTFTQFTYSSQSLVDPIVLKGL-TFTATGILTVVTQLPHH	240
Qy	260	SHGGRPEARLLVVVTDGESH-DGEELPAALKACEAGRVTTRYGIAVLGHLRRLRRDPSSF	318
Db	241	KNGARKSAKKILIVITDGQKYDPLEYSDVIPQAEKAGIRYAIQV-GHAF---QGPTA-	295
Qy	319	LREIRTTASDPDERFFNVTDAAALTDIVALGDRIRFLEGSHAENESSFGLEMSQIGFS	378
Db	296	RQELNTISSAPPQDPHVFKVDNFAALGSIQKLOEKIYAVEGTQGRASSSFFOHEMSQEGFS	355
Qy	379	THRLKDGILFGWCAYDWWGGSVLWLEGHRLFPFRMALEDFPPALQNH-----AY	430
Db	356	TALTMDGLFLGAVGSFW-----SGGAFILPPNMS-----PTFINMSQENVMDMSY	402
Qy	431	LGYSVSMLLRGRRRLFLSGAPRHRGKVIAF-QLKKDGAVRVAOQLQGQIGSIFYGSE	489

```

403 LGVSTELALWKGQVNLVL--GAPPYQHTGKAVITFTQVSRW--RRKAEVTCGTQSGVFGAS 459
490 LCPDLTDRTDGTDLVLLVAAPFMELPGONKETGRVYVYLV--GQSLTTLQCTLOQPEPPQD-- 546
460 LCSVDVDSGSTDLLILIGAPHY--BQTRGGQVSVCLPRGQVQVQCDLAVLKEGQHPW 517
547 ARPGFAMGALPDNLNQGFADVAVGAPLEDGHQCALYLYHG--TOSGVRPHPAQRIAAASMP 605
518 GREGAALTVLGDDVNEKLDLIDVAIGAPGEBQENRGAUVLFGHASESGISPHSHORIASQSLS 577
606 HALSYFGRSVDGRDLDDGDDLDVAVAGCAAAALLSSRPRIHVHTPSLLEVTPQAISVVVQD 665
578 PRLOYFGQALSSGQDLITQDGLMDLAVAGKQVLLRLSLPVLKVGVAWRSPVEVAKAVYR 637
666 C-----RRRGOEAVCLTAALCFQVTSRTPGRWDHQFYMRFTASLDEWTAGARAAFDGS 718
638 CWEKPSALEAGDATVCLTIQ-----KSSLDQLGDIOSSVRFDLALDPCLRTSRAIFNET 692
719 GQRLSPRRRLSVGNVTCBQLHPHVL D--TSDYLRPVALTVTFALDNTTRPG-----PVLN 772
693 KNPFTLRRKTKLGLG--IHCETLKLPLDCVEDVVSPIILHLNFSVLREPIPSQNLRLPVL A 751
773 EGSPTSIQKLVPFSCDGPNECVTDLVLQVNMDIRGSRKAPFVVRGRRKVLVSTTLEN 832
752 VGSQDLFTASLPPEKNCQDGLCEGD--LGVTLFSFGLQF---LTVGSSLELNVIVTVMN 806
833 RKENAYNTLSIIFSRNLHLASLTPORESF---IRVEC--AAPSAAHARL----CSVGHPV 883
807 AGEDSVGTVSVLYYPAGLSHRVRSVGAQKQPHQSALALACETVTEDEGLSSRCSVNHPI 866
884 FOTGAKVTFLLEPEFSCSLLSOVFKLTASSDSLRNGTLOBNTAQTSAYIOYEPHLLF 943
867 FHEGSGNTFIVTDFVSVKATLGRM--LMRASASSENKASSKATFOLELPVKYAVVTMI 925
944 SSESILHRYEVHYGTLVPVCPGPEFTKLRLVONLGCYVVSGLLIISALLPAVAGGNYFLS 1003
926 SROESTKY--FNFATSDKMKAEASHRYRVNLSQORDLA--ISINFVPVLLNGVAVMDV 982
1004 LSOVITNNASCIQNLTEPPGPPVHPPEELQHTNR---LNGSNTQCQVVRCHLQCLAKGTE 1060
983 VMEAPQSILPCVSR-----KPHQSDFLQTQISRSPMLDCSIADCLQFCDFVPSFSVOEE 1037
1061 VSVGLLRLVHNEFFRRAKFKSLTVVSTFELGTBEGSVLQLOTEASRWSESLLE--VWQTRPI 1119
1038 LDFTLKGNSLFGWVRETLQKKVLVWSVAEITFTDSVVSQLPQGEAFMRAQMEMVMEDEEV 1097
1120 LISLWILGSLGCLILLALLVFLCMLKGLGFF-AHKXIPIBEEKREE 1163
1098 YNAIPIIMSSVGALLLLALITATLYKLGFKRHYKEMLEDKPED 1142

RESULT 13
US-08-482-293A-2
; Sequence 2, Application US/08482293A
; Patent No. 5831029
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

RESULT 13

US-08-482-293A-2

; Sequence 2, Application US/08482293A

; Patent No. 5831029

; GENERAL INFORMATION:

APPLICANT: Gallatin, W. Michael

APPLICANT: Van der Vieren, Monica

; TITLE OF INVENTION: NO. 3831029ET Human z integrin Alpha subunit  
 ; NUMBER OF SEQUENCES: 103

; NUMBER OF SEQUENCES: 7  
: CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 233 South Wacker Drive, 6300 Sear Tower

; CITY: Chicago

; STATE: Illinois

COUNTRY: United States

ZIP: 60606-6402  
COMMITTEE READABLE FORM.

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:



```
/ APPLICATION NUMBER: US/08/482,293A
/ FILING DATE:
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/173,497
/ FILING DATE: 23-DEC-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/286,889
/ FILING DATE: 5-AUG-1994
/ APPLICATION DATA:
/ APPLICATION NUMBER: US 08/362,652
/ FILING DATE: 21-DEC-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Williams Jr., Joseph A.
/ REGISTRATION NUMBER: 38,659
/ REFERENCE/DOCKET NUMBER: 27866/32684
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 312-474-6300
/ TELEFAX: 312-474-0448
/ TELEX: 25-3856
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1161 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-482-293A-2

Query Match      18.58; Score 1131.5; DB 2; Length 1161;
Best Local Similarity 29.48; Pred. No. 1e-99;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

QY 23 FNLDEHHPLRPPGPEAEFGVSLQHVGGGORMWLVGAPWDPGSDRRGVDYRCPVGAH 82
DB 17 FNLDEVEPTIQ-EDAGGFGSVVQF--GGR-LVVGAPLEVVAANTGLDYC---AAA 69
QY 83 NAFCAKGLHDYQLGNSHSPAVNMHLGMSLLETDGDFMACAPLWBRACGSSVFSSGIC 142
DB 70 TGMCOPIPL-----HIRPEAVNMSLGLTLAASNTGSRLLACGPTLHRVCGENSYSKGSC 123
QY 143 ARVDASFOGSLAPTAQRCP-TYMDVVIVLDGNSI--YPMSEVQTFRLRLVQKLFIDP 199
DB 124 LLLGSRWEIITQTPDAPTECPHQEMDIVFLIDGSGSIDQNDFNQMGFVQAVMGQ--FEG 181
QY 200 EQIQVGLVQVGGESPVHWSLGDFTKEEVVRAAKNLSRREGRETKTAAQIMVACTEGFSQ 259
DB 182 TDTLPAIMQVSNLLKIHFTTQFTSPSQSLVDPIVQLKGL-TFTATGILTVVTQFLHH 240
QY 260 SHGRPEARLLVVVTDGESH-DGEELPAALKACEAGRVTYGIYAVLGHYLRQRDPSSF 318
DB 241 KNGARKSAKKILIVITDGQYKDPLEYSDDVIPQAEKAGIIRYAGV-GHAF---QGPTA- 295
QY 319 LREIRTIASDPDERFFNFVTDAAALTDIVDALGDRIFGLEGSHABNESSFGLEMSQIGFS 378
DB 296 ROELNTISSAPPQDHVFKVDNFALGSIQKLOQEKIYAVEGTOSRASSSFQHEMSQSGFS 355
QY 379 THRLKDGILFGWGVAGVWGGVSLWLEGGHRLFFPRMALEDEFPALQNH-----AY 430
DB 356 TALTMDELFLGAVGSFNS-----SGAFLYPPNMS-----PTFINMQENVDMRDSY 402
QY 431 LGYSVSMLLRGRRRLFLSGAPRFRHGKVIAP-QLKKDGAVRVAQSLQGEIQSGSYFGE 489
DB 403 LGYSTELALWGVQNLVL-GAPRYQHTGKAVIFTQVSRQW--RKAETVGTQIGSYFGAS 459
QY 490 LCPDLDTRDGTDTVLLVAAPMFLGPQNKETGRVVYLV--GQQSLTLQGTLPQPPQD- 546
DB 460 LCSVDVSDSDGSTDILIGAPHYY--EQTREGQGVSVCPPLPRGQVQWQCDVLRGEQGHWP 517
QY 547 ARFGFAGALPDNLQDGFADVAVGAPLEDGHQALYLVHG-TQSGVPHPAQRIATAASMP 605
DB 518 GRFAALTVLGDVNEKLDIVAGPGEQENRGAIVLFGHASESGISFSPHSQRIASSQLS 577
QY 606 HALSYFGRSVDGRDLDDLDGDDLDVAVGAQGAAILLSRRPIVHLTPSLEVTPQAISVVQRD 665

578 PRLOQFQALSGQDLTQDGLMDLAVGARGQVLLLRSLPLVKVGVAMRSPVEVAKAVYR 637
666 C-----RRRQGEAVCLTAALCFQVTSRTPGRWDOHQFYMRFTASIDETAGARAPDGS 718
638 CWEEKPSALEAGDATVCLTIQ-----KSLDQLGDIQSSVRFDLADPQLRSLRALFNET 692
719 GQRLSPRRRLRLSVGNVTCBQLHFPVLD-TSDYLRLPVALTFTFALONTTKPG-----PVLN 772
693 KNPTLTKRTKLG-LHCETLKLKLLPDCVEDVVSPIILHLNLSLVREPSPQNLRLPVLA 751
773 EGSPTSIQKLVPFSGKDCPDNECVTDLVLOVNMIDIRGSRKAPVVRGRRKVLVSTLEN 832
752 VGSQDLFTASLPKNCQGDGLCEGD--LGVTLFSGLQT--LTVGSSLELNLVITVWN 806
833 RKENAYNTSLSIIFSRNLHLASLTPORESP-----IKVEC-AAPSAHARL-----CSVGHVP 883
807 AGEDSYGTVVVSLYYPAGLSHRRVSGAQKQPHOSALRLACETVPTDEGLRSSKCSVNHP 866
884 FOTGAKVTFLLEPEFSCSLLSQVFGKLTASSDSLERNGTLOENTTAQTSAIYQYEPHLLP 943
867 FHGSGNGTIVTFDVSYSKATLGDRLM-LMRASASSENKASSKATFQLELPVKYAVYVMI 925
944 SSESTLHREYVHPYGTLPVGPGEFKTTLRVQNLGYVVSGLIISALLPAVAHGGNYFLS 1003
926 SRQESTKY--FNFATSDKMKKEAHRVYRVNLSQORDLA-ISINFWVPVLLGVAVMDV 982
1004 LSQVITNNAASCIVONLTPEPPVHPPEELOHNR---LNGSNTQCQVVRCHLQOLAKGTE 1060
983 VMRAPOSQSLPCVSR-----KPPQHSDFLTQISRSPMLDCSIADCLQFCDVPSFSVQEE 1037
1061 VSVGLLRLVHNEPFRRAKFKSLTVVSTFELGTBEGSVLQLTEASRWSESLLE-VVQTRPI 1119
1038 LDTLKGNSLPGWVRITLQKVLVSVVAEITPDTSVYSLQPCQEAFFRAQEMVLESEDEV 1097
1120 LISLWILGSLVGLLALLLALLVCLMKLQFF-AHKKIPEEKREE 1163
1098 YNAIPIIMSSVVGALLLALITATLYKLGFFKGRHYKEMLEDKPED 1142

RESULT 14
US-08-943-363-2
; Sequence 2, Application US/08943363
; Patent No. 5837478
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,363
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
```

[illegible]

Db 182 TDTLFAIMOYNLLKHIFTTQFRTSPSQSLVDPIVOLKGL-TFTATGILTUVVTOLFHH 240  
Qy 260 SHGGRPEAAALLVVVTDGESH-DGEBULPAALKAACEAGRVTRYGIAVLGHVYLRQRDPSSP 318  
Db 241 KNGARKSAKILIVITDGGYKDPLEYSVDVPOEKAGIIRYALGV-GHAF---QGPTA- 295  
Qy 319 LREIRTIASDPDERFFNFVDEAALTDIVDALGDRIFGLEGSHAENESSFGLMSQIGFS 378  
Db 296 RQELNTISSAPPQDHVFKVDFNFAALGSIQKOLQEKIYAVSGTQSRASSSFQHEMSQEGFS 355  
Qy 379 THRLKQILFQMGVAYDWGGSVLMLEGGHRLFPPRMALEDEFPFALQNH-AY 430  
Db 356 TALTMGDLFLGAVGSFSW-----SGGAFLYPNMS-----PTFINMSQENVDMRDSY 402  
Qy 431 LGYSVSMLLRGGRRLFLGAPRHRGKVIAP-OLAKDGAVRVAQSLQBGQIGSYFGSE 489  
Db 403 LGYSTELALWKGVOQLVL-GAPRYQHTGKAVIFTQVSRQW--RKAEVGTGTQIGSYFGAS 459  
Qy 490 LCPLDTRDGTDTLLVAAAMFLGPONKETGRVYVYL- -GQOSLLTLQGTLOPEPPQD- 546  
Db 460 LCSVDVDSGSTDLILIGAPHY--EQTRGQVSVCPPLRGQRVQWQCDVLRGEGQHPW 517  
Qy 547 ARFGFANGALPDLNQDGFADVAGAPLEDGHQGALYLYHG-TQSGVRPHPAQRIAAASMP 605  
Db 518 GRFGAALTVLGDVNNEDKLIDVAIGAPGEQENRGAVILFHGASESGISPSHSQRIASSQLS 577  
Qy 606 HALSYFGRSVDGRDLDDGDDLDVAVGAGAAILLSSRPVHLTPSLVTPFQALSVVQRD 665  
Db 578 PRLOQYFQALSGGGQDLTQDGLMDLAVGARGQVLLRLSLPLKVGAMRFPSPVEVAKAYR 637  
Qy 666 C-----RRRGQBAVCLTAALCFQVTSRTPTGRMDHQFYMRFTASLDDEWTAGARAAFDGS 718  
Db 638 CWEKPSALEAGDATVCLTIQ-----KSSLDQLGDIQSSVRFDLALDPGRLTGRAIENET 692  
Qy 719 GQRLSPRLRLSVGNVTCBQLHFHVL-DSDYLRPVALTVTFALDNTTKPG-----PVLN 772  
Db 693 KNPTLTRRKTILGLG-IHCETILKLLPDCVEDVVSPIILHLNFSLVREPIPSQNLRPVLA 751  
Qy 773 EGSPSTSIQKLVPRSKDGPNECVTDVLQVNMDIRGSRKAPFVVRGRRKVLVSTTLEN 832  
Db 752 VGSQDLFTASLPFKNCGQGLCEGD--LGVTLSFGLQT---LTVGSSLELNVIVTWN 806  
Qy 833 RKENAYNTSLSIIFSRNLHLASLTPQRESP----IKVEC-AAPSAHARL----CSVGHPV 883  
Db 807 AGEDSYGTVVSLYYYPAGLSHRRVSGAQKQPHQSALRLACETVPTDEGLSSRCVNHPI 866  
Qy 884 PQTGAKVTFLLEPFSCSSLLSQVFGKLTASSDSLERNGTLQENTAQTSAYIQYEPHLLF 943  
Db 867 PHEGNGTFTVTPDVSYKATIGDRM-LMRASASSENKASSKATFQLELPVKYAVYTM 925  
Qy 944 SSESTLHRYEVHPYGTLPVGPPEFKTLRVONLGCYVVSGLIISALLPAVAHGNYPFLS 1003  
Db 926 SRQESTKY--FNFATSEKOKKEAHRVNNLSQDRLA-ISINFVVPVLLNGVAVDV 982  
Qy 1004 LSQVITNNASCIVQNLTEPPGPPVHPBELQHTNR---LNGSNTQCQVVRCHLQLAGTE 1060  
Db 983 VMEAPSQLPCVSE- -KPPQHSDFLTQISRSPMLDCSIADCLQFRCDVPSPSVQEE 1037  
Qy 1061 VSVGLLRVLVNHNEFFRRAKFSKLTVVSTFELGTGEGSVLQLTEASRWSESLE-VVQTRPI 1119  
Db 1038 LDFTLKGNLSFGVRETILQKKVLVSVVAEITFDTSVYSQLPQGBAFWRAQMEMVLEBDEV 1097  
Qy 1120 LISLWILIGSVLGLLIALVCLMKLGP-AHKKIPEEKREE 1163  
Db 1098 YNAIPIIMGSSVGALLLALITATLYLXGFFRHYKMWLEDKPED 1142

Search completed: April 6, 2005, 12:24:20  
Job time : 54.8054 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 6.89976 Seconds  
(without alignments)  
2761.097 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337

Perfect score: 1025

Sequence: 1 GICARVDASFPQGSGLAPTA.....FLREIRTIASDPDERFFNV 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.\*

2: Pir2.\*

3: Pir3.\*

4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	473	46.1	272	2 A55348	integrin alpha-1 -
2	467.5	45.6	1151	2 A45226	integrin alpha-1 c
3	465.5	45.4	1180	2 A35854	integrin alpha-1 c
4	454	44.3	1178	2 S44142	VLA-2 protein homo
5	451	44.0	1170	2 I45914	integrin alpha 2 s
6	441	43.0	1181	2 A33998	integrin alpha-2 c
7	336	32.8	191	2 I47230	VLA-2 protein - pi
8	227.5	22.2	1747	2 A45974	collagen alpha 1(X
9	227.5	22.2	1857	2 S31212	collagen alpha 1(X
10	227.5	22.2	1888	2 S78476	collagen alpha 1(X
11	227.5	22.2	3124	2 A40020	collagen alpha 1(X
12	214	20.9	493	2 A33809	cartilage matrix p
13	205	20.0	496	2 A37979	cartilage matrix p
14	204.5	20.0	929	2 I51027	type XII collagen
15	201.5	19.7	3137	2 A37797	collagen alpha 3(V
16	199	19.4	3051	2 S42373	hypothetical prote
17	194.5	19.0	500	2 S65522	cartilage matrix p
18	177.5	17.3	1179	2 A53213	integrin alpha-E c
19	176.5	17.2	3176	2 CGH3A	collagen alpha 3(V
20	175	17.1	724	2 A48569	antigen Em100 - Ei
21	170.5	16.6	843	2 A40970	undulin 1 - human
22	169	16.5	741	2 T46488	hypothetical prote
23	158.5	15.5	1153	1 RWHU8	cell surface glyco
24	158	15.4	1163	2 I56126	lymphocyte fuction
25	157.5	15.4	1153	2 S0051	leukocyte surface
26	156	15.2	550	2 T23760	hypothetical prote
27	156	15.2	2944	2 A54849	collagen alpha 1(V
28	154	15.0	2813	1 VWHU	von Willebrand fac
29	153.5	15.0	1163	1 RWHUIC	cell surface glyco

30	152	14.8	1170	2 S03308	cell surface glyco
31	142.5	13.9	712	2 A45638	immunodominant mic
32	141	13.8	341	2 T32949	hypothetical prote
33	136.5	13.3	427	2 G00039	von Willebrand fac
34	129.5	12.6	414	2 PS0323	protein M01E10.2 (
35	123	12.0	1286	2 A88396	hypothetical prote
36	116.5	11.4	13055	2 T16580	hypothetical prote
37	114.5	11.2	567	2 T28797	hypothetical prote
38	112.5	11.0	661	2 T16597	hypothetical prote
39	110	10.7	643	2 T19549	hypothetical prote
40	107	10.4	449	2 T33257	protein CTRP - mal
41	103.5	10.1	2098	2 T18397	collagen alpha 2(V
42	100	9.8	1029	1 S21369	collagen alpha 1(V
43	97.5	9.5	780	2 A34102	von Willebrand fac
44	95.5	9.3	1019	1 A32856	undulin 2 - human
45	95	9.3	445	2 B40970	

ALIGNMENTS

RESULT 1

A55348

integrin alpha-1 - chicken (fragment)

C:Species: Gallus gallus (chicken)

C>Date: 10-Feb-1995 #sequence\_revision 10-Feb-1995 #text\_change 09-Jul-2004

C:Accession: A55348

R;Kern, A.; Briesewitz, R.; Bank, I.; Marcantonio, E.E.

J. Biol. Chem. 269, 22811-22816, 1994

A:Title: The role of the I domain in ligand binding of the human integrin alpha-1beta-1.

A:Reference number: A55348; MUID:94357930; PMID:7521332

A:Accession: A55348

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-272 <KER>

A:Cross-references: UNIPROT:O42094; GB:U10114

F;55-230/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 46.1%; Score 473; DB 2; Length 272;

Best Local Similarity 45.5%; Pred. No. 7.7e-36;

Matches 90; Conservative 37; Mismatches 71; Indels 0; Gaps 0;

Qy	1	GICARVDASFPQGSGLAPTAQRCPYMDVIVLVDGNSIYPWSEVQVFLRELVLGKLFDP	60
Db	30	GVCSNVSSTFETVKAVAPSVQVECKTQDIDIVLVDGNSIYPWSEVTAFLNLLRNMDIGP	89
Qy	61	EQIQVGLVQYGESPVHEWSLGDPRTEKEVRAAKNLSRREGRETQTAQINVACTEGFSQ	120
Db	90	QQTQVGIVQYQGVVHEFVLTNTYSTTEEVMDAALRIQRGGTQTWTALGIDTAREEAFTE	149
Qy	121	SHGGRPEARLLVVVTVDGSHDGBELPAALKACAGRVTRYGVIAVLGHYLRQRDPSSFL	180
Db	150	AHGARRGVQVMVITVDGSHDNYRLQEVIDKCEDENIQRFPAIILGYSRGNLSTKFFV	209
Qy	181	REIRTIASDPDERFFNV 198	
Db	210	EIIKSIASKPTKHFNV 227	

RESULT 2

A45226

integrin alpha-1 chain - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 30-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004

C:Accession: A45226

R;Briesewitz, R.; Epstein, M.R.; Marcantonio, E.E.

J. Biol. Chem. 269, 2989-2996, 1993

A:Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.

A:Reference number: A45226; MUID:93155124; PMID:8428973

A:Accession: A45226

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1151 <BRI>

A;Cross-references: UNIPROT:P56199  
A;Experimental source: Hepatocarcinoma cell line HepG2  
A;Note: sequence extracted from NCBI backbone (NCBIP:124326)  
F;142-317/Domain: von Willebrand factor type A repeat homology <VWA1>

Query Match 45.6%; Score 467.5; DB 2; Length 1151;  
Best Local Similarity 46.5%; Pred. No. 1.4e-34;  
Matches 92; Conservative 36; Mismatches 69; Indels 1; Gaps 1;

Qy 1 GICARVDASFOQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFRLRLVGLKFLIDP 60  
Db 118 GICSDVSPDFQVNSIAP-VQECSTQLDIVLDGNSIYPWSDSVTAFLNLLKRMIDGP 176  
Qy 61 EQIQVGLVQVGESEPVHWSLGDFTKBEVVAANKLSRRREGRETKTAQAIMVACTEFSQ 120  
Db 177 KQTQVGIQVGENVTHFNLNKYSTBEVLVAANKIVQGRGQRTWTALGDTTARKEAFTE 236  
Qy 121 SHGGRPEAARLLVVVTDGESHGDEELPAALKACBAGRVTRYGIAVLGHYLRQRDPSSFL 180  
Db 237 ARGARRGVKVMVITVDGESHGDEELPAALKACBAGRVTRYGIAVLGHYLRQRDPSSFL 180  
Qy 181 REIRTIASDPDERFFNV 198  
Db 297 EEIKSIASEPTEKHFFNV 314

RESULT 3  
A35854  
integrin alpha-1 chain precursor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 23-Oct-1990 #sequence\_revision 13-Sep-1991 #text\_change 09-Jul-2004  
C;Accession: A35854; S11243  
R;Ignatius, M.J.; Large, T.H.; Houde, M.; Tawil, J.W.; Barton, A.; Esch, F.; Carbonetto,  
J.; Cell Biol. 111, 709-720, 1990  
A;Title: Molecular cloning of the rat integrin alpha-1 subunit: a receptor for laminin a  
A;Reference number: A35854; MUID:90338125; PMID:2380249  
A;Accession: A35854  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-1180 <IGN>  
A;Cross-references: UNIPROT:P18614; GB:X52140; NID:956493; PIDN:CAA36384.1; PID:956494  
A;Keywords: cell adhesion; cytoskeleton; transmembrane protein  
F;170-345/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 45.4%; Score 465.5; DB 2; Length 1180;  
Best Local Similarity 46.5%; Pred. No. 2.2e-34;  
Matches 92; Conservative 35; Mismatches 70; Indels 1; Gaps 1;

Qy 1 GICARVDASFOQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFRLRLVGLKFLIDP 60  
Db 146 GICSDVSPDFQVNSIAP-VQECSTQLDIVLDGNSIYPWSEVQTFRLRLVGLKFLIDP 204  
Qy 61 EQIQVGLVQVGESEPVHWSLGDFTKBEVVAANKLSRRREGRETKTAQAIMVACTEFSQ 120  
Db 205 KQTQVGIQVGENVTHFNLNKYSTBEVLVAANKIGRQGLQRTWTALGDTTARKEAFTE 264  
Qy 121 SHGGRPEAARLLVVVTDGESHGDEELPAALKACBAGRVTRYGIAVLGHYLRQRDPSSFL 180  
Db 265 ARGARRGVKVMVITVDGESHGDEELPAALKACBAGRVTRYGIAVLGHYLRQRDPSSFL 180  
Qy 181 REIRTIASDPDERFFNV 198  
Db 325 EEIKSIASEPTEKHFFNV 342

RESULT 4  
S44142  
VLA-2 protein homolog - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004  
C;Accession: S44142  
R;Edelman, J.M.; Chan, B.M.; Uniyal, S.; Onodera, H.; Wang, D.Z.; Damjanovich, L.; Latze  
; submitted to the EMBL Data Library, January 1994

A;Description: The mouse VLA-2 homologue supports collagen and laminin adhesion but not  
A;Reference number: S44142  
A;Accession: S44142  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-1178 <EDE>  
A;Cross-references: UNIPROT:Q62469; EMBL:Z29987; NID:9473098; PIDN:CAA82877.1; PID:947309  
F;169-344/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 44.3%; Score 454; DB 2; Length 1178;  
Best Local Similarity 46.0%; Pred. No. 2.5e-33;  
Matches 91; Conservative 33; Mismatches 74; Indels 0; Gaps 0;

Qy 1 GICARVDASFOQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFRLRLVGLKFLIDP 60  
Db 144 GICSDVSPDFQVNSIAP-VQECSTQLDIVLDGNSIYPWSEVQTFRLRLVGLKFLIDP 203  
Qy 61 EQIQVGLVQVGESEPVHWSLGDFTKBEVVAANKLSRRREGRETKTAQAIMVACTEFSQ 120  
Db 204 KKTQVALIQVANEPRIFNLNDPETKEDMVQATSETRQHGGLTNTFRAIEFARDVAYSQ 263  
Qy 121 SHGGRPEAARLLVVVTDGESHGDEELPAALKACBAGRVTRYGIAVLGHYLRQRDPSSFL 180  
Db 264 TSGRPGATKVMVITVDGESHGSKLTKVIOCCNDDEILRFGIAVLGYNLRNALDTKNLI 323  
Qy 181 REIRTIASDPDERFFNV 198  
Db 324 KEIKAIASIPTEKHFFNV 341

RESULT 5  
I45914  
integrin alpha 2 subunit - bovine (fragment)  
C;Species: Bos primigenius taurus (cattle)  
C;Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 09-Jul-2004  
C;Accession: I45914  
R;Kamata, T.; Puzon, W.; Takada, Y.  
J. Biol. Chem. 269, 9659-9663, 1994  
A;Title: Identification of putative ligand binding sites within the I-domain of integrin  
A;Reference number: A54402; MUID:94193647; PMID:7511592  
A;Accession: I45914  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-1170 <KAM>  
A;Cross-references: UNIPROT:P53710; GB:L25886; NID:9439695; PIDN:AAB59255.1; PID:9439695  
F;161-336/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 44.0%; Score 451; DB 2; Length 1170;  
Best Local Similarity 44.4%; Pred. No. 4.6e-33;  
Matches 88; Conservative 37; Mismatches 73; Indels 0; Gaps 0;

Qy 1 GICARVDASFOQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFRLRLVGLKFLIDP 60  
Db 136 GICSDVSPDFQVNSIAP-VQECSTQLDIVLDGNSIYPWSEVQTFRLRLVGLKFLIDP 195  
Qy 61 EQIQVGLVQVGESEPVHWSLGDFTKBEVVAANKLSRRREGRETKTAQAIMVACTEFSQ 120  
Db 196 TKTQVGLIQVANNPRVFNLTFSKDEMIKATSTQFYQGDLTNTFKAIQYARDTAYST 255  
Qy 121 SHGGRPEAARLLVVVTDGESHGDEELPAALKACBAGRVTRYGIAVLGHYLRQRDPSSFL 180  
Db 256 AAGGRPGATKVMVITVDGESHGSKLKAVIDQCCKNILRFGIAVLGYNLRNALDTKNLI 315  
Qy 181 REIRTIASDPDERFFNV 198  
Db 316 KEIKAIASIPTEKHFFNV 333

RESULT 6  
A33998  
integrin alpha-2 chain precursor - human  
N;Alternate names: CD49b; platelet glycoprotein GPIa; VLA-2/collagen receptor alpha-2 ch  
C;Species: Homo sapiens (man)

A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-191 <BAH>  
A;Cross-references: UNIPROT:Q29124; EMBL:Z12137; NID:g2158; PIDN:CAA78125.1; PID:g2159  
C;Keywords: glycoprotein

Query Match 32.8%; Score 336; DB 2; Length 191;  
Best Local Similarity 43.9%; Pred. No. 1.9e-23;  
Matches 69; Conservative 27; Mismatches 61; Indels 0; Gaps 0;

QY 42 WSEVQVFLRLVGLKFLIDPEIOIGVLVOYGESPVHWSLGDPRTKKEEVVRAAKNLSRREG 101  
DB 1 WDAKNFLFKFVGLGLDGPTKTVGLIQANNPRVFNLTFTKKAEMWEATSHTTQYGG 60  
QY 102 RETTQAQIMVACTEGFSOSHGRPEARLLVVVTVDGESHGDELPALAKACAGRVTRY 161  
DB 61 DLNTFTKAIQAYRDSAYSAAAGRGPGATKVMVVVTVTGDESHGSLKXAVIDQCNDNLRF 120  
QY 162 GIAVLGHYLRQRDPSSFLREIRTIASDDPERFFNV 198  
DB 121 GIAVLGYLRNALDTKNLIKEIKAIASIPTRYFFNV 157

RESULT 8  
A45974  
collagen alpha 1(XIV) chain precursor, short form 2 - chicken  
N;Alternate names: undulin  
C;Species: Gallus gallus (chicken)  
C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 09-Jul-2004  
C;Accession: A45974; S30085; S22916; S17035; S20833  
R;Gerecke, D.R.; Foley, J.W.; Castagnola, P.; Gennari, M.; Dublet, B.; Cancedda, R.; Lin  
J. Biol. Chem. 268, 12177-12184, 1993  
A;Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' region  
ns.  
A;Reference number: A45974; MUID:93280195; PMID:8505337  
A;Accession: A45974  
A;Status: preliminary  
A;Molecule type: mRNA; protein  
A;Residues: 1-1747 <GER>  
A;Cross-references: UNIPROT:P32018  
A;Experimental source: embryo skin  
A;Note: sequence inconsistent with the nucleotide translation  
A;Note: sequence extracted from NCBI backbone (NCBIN:133364, NCBIP:133365)  
R;Apte, S.S.  
submitted to the EMBL Data Library, March 1992  
A;Reference number: S30085  
A;Accession: S30085  
A;Molecule type: mRNA  
A;Residues: 1472-1660 <APT>  
A;Cross-references: EMBL:X65122; NID:g62871; PIDN:CAA46238.1; PID:g938175  
R;Trueb, J.; Trueb, B.  
Eur. J. Biochem. 207, 549-557, 1992  
A;Title: Type XIV collagen is a variant of undulin.  
A;Reference number: S22916; MUID:92339443; PMID:1339349  
A;Accession: S22916  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 286-494, 'Q', 496-834, 'A', 836-1119, 'XL', 1122-1402, 1409-1439 <TRU>  
R;Gordon, M.K.; Castagnola, P.; Dublet, B.; Linsemayer, T.F.; van der Rest, M.; Mayne,  
Eur. J. Biochem. 201, 333-338, 1991  
A;Title: Cloning of a cDNA for a new member of the class of fibril-associated collagens  
A;Reference number: S17035; MUID:92037585; PMID:1935930  
A;Accession: S17035  
A;Molecule type: mRNA  
A;Residues: 1472-1659 <GOR1>  
A;Accession: S20833  
A;Molecule type: protein  
A;Residues: 1551-1570, 1593-1599, 1639-1667 <GOR2>  
C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer  
F;40-204/Domain: von Willebrand factor type A repeat homology <VWA>  
F;236-317/Domain: fibronectin type III repeat homology <FN3A>  
F;326-409/Domain: fibronectin type III repeat homology <FN3B>  
F;418-498/Domain: fibronectin type III repeat homology <FN3C>





F;995-1076/Domain: fibronectin type III repeat homology <FN3G>  
F;1086-1169/Domain: fibronectin type III repeat homology <FN3H>  
F;1197-1361/Domain: von Willebrand factor type A repeat homology <VWA3>  
F;1384-2295/Domain: IID #status predicted <IID>  
F;1384-1465/Domain: fibronectin type III repeat homology <FN3I>  
F;1474-1557/Domain: fibronectin type III repeat homology <FN3J>  
F;1566-1647/Domain: fibronectin type III repeat homology <FN3K>  
F;1655-1738/Domain: fibronectin type III repeat homology <FN3L>  
F;1756-1838/Domain: fibronectin type III repeat homology <FN3M>  
F;1847-1928/Domain: fibronectin type III repeat homology <FN3N>  
F;1937-2019/Domain: fibronectin type III repeat homology <FN3O>  
F;2028-2110/Domain: fibronectin type III repeat homology <FN3P>  
F;2119-2199/Domain: fibronectin type III repeat homology <FN3Q>  
F;2207-2294/Domain: fibronectin type III repeat homology <FN3R>  
F;2325-2490/Domain: von Willebrand factor type A repeat homology <VWA4>  
F;2438-2440/Region: cell adhesion #status predicted  
F;2509-2750/Domain: IXP, homologous to NC4 domain of type IX collagen #status predicted  
F;2751-2902/Domain: collagenous COL2 #status predicted <COL2>  
F;2899-2901/Region: cell attachment (R-G-D) motif  
F;2903-2945/Domain: non-collagenous NC2 #status predicted <NC2>  
F;2946-3048/Domain: collagenous COL1 #status predicted <COL1>  
F;3049-3124/Domain: non-collagenous NC1 #status predicted <NC1>  
F;32, 1006.1032, 1044, 1512, 1767, 2210, 2273, 2532, 2683/Binding site: carboxyproline (Asn)  
F;2780, 2789, 2836, 2842, 2860, 2866, 2869, 3004, 3007/Modified site: hydroxyproline (Pro) #sta

Query Match 22.2%; Score 227.5; DB 2; Length 3124;  
Best Local Similarity 34.1%; Pred. No. 4.9e-12;  
Matches 59; Conservative 32; Mismatches 67; Indels 15; Gaps 5;

QY 28 DVIVILDGNSI--YPWSEVOTFLRLVLKGLFDIPQIQVLGVQYESPVHEWSLGDFRT 85  
:::|||||::: |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
DB 1199 DILLVDGWSIGRPNFKNFRISRIVEFDIPGPQVIGLAQYGDPRTENLNAYRT 1258  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
QY 86 KEKVRAAKNLRRREGRETQAQIAIWACTEGFSQHSGRPEARLLVVVTDCSHDGE 145  
||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
DB 1259 KEALLDAVNLPVK-GENTUTGMALDPILKNFKQAGLRPRARKIUGLVITDKGSQDDVV 1317  
||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
QY 146 LPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDERFFPNV 198  
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::  
DB 1318 TPSERLRDEG-VELVAIGI-----KNADN---ELKQIATDPDDIHAYNV 1358  
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::

RESULT 12  
A33809  
Cartilage matrix protein precursor - chicken  
C:Species: Gallus gallus (chicken)  
C>Date: 09-Mar-1990 #sequence\_revision 09-Mar-1990 #text\_change 09-Jul-2004  
C:Accession: A33809; A26364  
R:Kiss, I.; Deak, F.; Holloway Jr., R.G.; Delius, H.; Mebust, K.A.; Frimberger, E.; Argraves, W.S.; Deak, F.; Sparks, K.J.; Kiss, I.; Goetinck, P.F.  
J. Biol. Chem. 264, 8126-8134, 1989  
A:Title: Structure of the gene for cartilage matrix protein, a modular protein of the extracellular matrix, von Willebrand factor, complement factors B and C2, and epidermal growth factor.  
A:Reference number: A33809; MUID:89255246; PMID:2542265  
A:Accession: A33809  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-493 <ARG>  
A:Cross-references: UNIPROT:P05099; GB:X12346; GB:X12347; GB:X12348; GB:X12349; GB:X12350  
R:Argraves, W.S.; Deak, F.; Sparks, K.J.; Kiss, I.; Goetinck, P.F.  
Proc. Natl. Acad. Sci. U.S.A. 84, 464-468, 1987  
A:Title: Structural features of cartilage matrix protein deduced from cDNA.  
A:Reference number: A26364; MUID:87092429; PMID:3025875  
A:Accession: A26364  
A:Molecule type: mRNA  
A:Residues: 78-493 <ARG>  
A:Cross-references: GB:M14792; NID:g211545; PID:AAA48695.1; PID:g211546  
C:Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repeat homology <VWA1>  
F;37-204/Domain: von Willebrand factor type A repeat homology <VWA1>  
F;225-260/Domain: EGF homology <EGF>  
F;270-434/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 20.9%; Score 214; DB 2; Length 493;  
Best Local Similarity 33.3%; Pred. No. 9.3e-12;



F:642-807/Domain: von Willebrand factor type A repeat homology <VW04>  
F:840-1004/Domain: von Willebrand factor type A repeat homology <VW05>  
F:1033-1197/Domain: von Willebrand factor type A repeat homology <VW06>  
F:1237-1400/Domain: von Willebrand factor type A repeat homology <VW07>  
F:1439-1604/Domain: von Willebrand factor type A repeat homology <VW08>  
F:1639-1804/Domain: von Willebrand factor type A repeat homology <VW09>  
F:1838-2010/Domain: von Willebrand factor type A repeat homology <VW10>  
F:2043-2378/Domain: collagenous #status predicted <COL>  
F:2045-2047/Region: cell attachment (R-G-D) motif  
F:2153-2155/Region: cell attachment (R-G-D) motif  
F:2159-2161/Region: cell attachment (R-G-D) motif  
F:2379-3137/Domain: non-collagenous #status predicted <CNC>  
F:2405-2577/Domain: von Willebrand factor type A repeat homology <VW11>  
F:2623-2806/Domain: von Willebrand factor type A repeat homology <VW12>  
F:2803-2846/Domain: platelet glycoprotein IB-like #status predicted <GPI>  
F:2954-3039/Domain: fibronectin type III repeat homology <FN3>  
F:3072-3122/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>  
F:201,2084,2436,2563,2581,2683,2867,2920,3003/Binding site: carbohydrate (Asn) (covalent)

Query Match 19.7%; Score 201.5; DB 2; Length 3137;  
Best Local Similarity 29.8%; Pred. No. 1.2e-09;  
Matches 59; Conservative 42; Mismatches 72; Indels 25; Gaps 6;  
Qy 2 ICARVDASFOQGS LAP-----TAQCPTVMDVVIVLDGNSI--YPSVSVQTFRLRLV 53  
Db 212 LVASVRTSMTPQAGAKGLVKDITAQES---ADLIFLDGSDNIGSVNFQAIRDFLVNLI 268  
Qy 54 GKLFIDPEIQVGLVQYGESPVHWSLGDPTKKEEVVRAAKNLSRRREGRETAKAIVVA 113  
Db 269 ESLRVGQAQIHIGVVQYSDQPRTEFALNSYSTKADVLDAVKALSRFGKEANTGALEYV 328  
Qy 114 CTEGFSQSHGGRPEAA--RLIVVVTGDGSHGELPAALKACEAGRTRYGIAVLGHVLR 171  
Db 329 VENLFTQAGGSRIEBAVPQILVLISGGSSD--DIREGLAVKQASIFSISIGVL----- 381  
Qy 172 RQDPSSFLREIRTIASD 189  
Db 382 -----NADSAELQIATD 394

Search completed: April 6, 2005, 12:17:03  
Job time : 8.89977 secs

***This Page Blank (usp10)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 29.3046 Seconds  
(without alignments)  
3459.921 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337  
Perfect score: 1025  
Sequence: 1 GICARVDASFOQGS LAPTA.....FLREIRTIASDPDERFFNV 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1025	100.0	1167	1 ITAG_HUMAN	O75578 homo sapien
2	547	53.4	1188	2 Q7QC3	Q7qc3 mus musculus
3	544	53.1	1189	1 ITAH_HUMAN	Q9ukx5 homo sapien
4	543	53.0	1188	1 ITAH_MOUSE	P61622 mus musculus
5	473	46.1	285	1 ITAI_CHICK	Q90615 gallus gall
6	473	46.1	1171	2 Q42094	O42094 gallus gall
7	467.5	45.6	1151	1 ITAI_HUMAN	P56199 homo sapien
8	465.5	45.4	1180	1 ITAI_RAT	P18614 rattus norv
9	454	44.3	1178	1 ITA2_MOUSE	Q62469 mus musculus
10	454	44.3	1178	2 Q6P1C7	Q6p1c7 mus musculus
11	451	44.0	1170	1 ITA2_BOVIN	P53710 bos taurus
12	441	43.0	1181	1 ITA2_HUMAN	P17301 homo sapien
13	389	38.0	323	2 Q8CB84	Q8cb84 mus musculus
14	336	32.8	191	2 Q29124	Q29124 sus scrofa
15	299	29.2	156	2 Q71V33	Q71v33 homo sapien
16	227.5	22.2	1888	1 CA1E_CHICK	P32018 gallus gall
17	227.5	22.2	3124	1 CA1C_CHICK	P13944 gallus gall
18	214.5	20.9	1797	2 Q80X19	Q80x19 mus musculus
19	214	20.9	493	1 Q80X19	P05099 gallus gall
20	213.5	20.8	3119	1 CA1C_MOUSE	Q60847 mus musculus
21	209.5	20.4	3063	1 CA1C_HUMAN	Q99715 homo sapien
22	209	20.4	637	2 Q81VX1	Q81vx1 homo sapien
23	209	20.4	1284	2 Q6PI59	Q6pi59 homo sapien
24	207.5	20.2	1472	2 Q90ZAO	Q90zao gallus gall
25	207	20.2	821	2 Q6PYX2	Q6pyx2 brachydanio
26	205.5	20.0	1329	1 KF10_HUMAN	Q9P218 homo sapien
27	205	20.0	496	1 CAMA_HUMAN	P21941 homo sapien
28	204.5	20.0	929	1 CA1C_NOTVI	Q91145 notophthalm
29	203.5	19.9	500	2 Q80VN5	Q80vn5 mus musculus
30	203.5	19.9	1816	2 Q8NT15	Q8nt15 brachydanio
31	201.5	19.7	3137	1 CA36_CHICK	P15989 gallus gall

#### RESULT 1

ID	ITAG_HUMAN	STANDARD;	PRT;	1167 AA.
AC	O75578; Q9UH28;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	05-JUL-2004 (Rel. 44, Last annotation update)			
DE	Integrin alpha-10 precursor.			
GN	Name=ITGA10;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Articular chondrocytes;			
RX	MEDLINE=98352078; PubMed=9685391; DOI=10.1074/jbc.273.32.20383;			
RA	Camper L., Hellman U., Lundgren-Aakerlund E.,			
RT	"Isolation, cloning, and sequence analysis of the integrin subunit alpha10, a beta1-associated collagen binding integrin expressed on chondrocytes.";			
RL	J. Biol. Chem. 273:20383-20389(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Endothelial cells, and Heart;			
RX	MEDLINE=20169197; PubMed=10702680;			
RA	Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,			
RA	Wang S.-X., Langley R., Krissansen G.W.;			
RT	"The integrin alpha10 subunit: expression pattern, partial gene structure, and chromosomal localization.";			
RL	Cytogenet. Cell Genet. 87:238-244(1999).			
CC	-!- FUNCTION: Integrin alpha-10/beta-1 is a receptor for collagen.			
CC	-!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-10 associates with beta-1.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-!- TISSUE SPECIFICITY: Widely expressed with highest expression in muscle and heart. Found in articular cartilage.			
CC	-!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins with I-domains do not undergo protease cleavage.			
CC	-!- SIMILARITY: Belongs to the integrin alpha chain family.			
CC	-!- SIMILARITY: Contains 7 FG-GAP repeats.			
CC	-!- SIMILARITY: Contains 1 VWFA domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.iesb-sib.ch/announce/ or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL; AF074015; AAC31952.1; -			
DR	EMBL; AF112345; AAP21944.1; -			
DR	EMBL; AF172723; AAP61638.1; -			

32	201	19.6	490	2	O7SVT5	O7svt5 xenopus lae
33	199	19.4	3183	2	Q652C2	Q652c2 caenorhabdi
34	199	19.4	3767	1	MUA3_CABEL	P34576 caenorhabdi
35	197	19.2	810	2	Q8C435	Q8c435 mus musculu
36	196	19.1	956	1	MTN2_MOUSE	O08746 mus musculu
37	196	19.1	956	2	Q8R542	Q8r542 mus musculu
38	195.5	19.1	422	2	Q8K1I8	Q8k1i8 dipodomys m
39	194.5	19.0	500	1	CAMA_MOUSE	P51942 mus musculu
40	193.5	18.9	1160	2	Q8MKF4	Q8mkf4 feils silve
41	193	18.8	451	2	Q8N2G3	Q8n2g3 homo sapien
42	191	18.6	280	2	Q8C4R9	Q8c4r9 mus musculu
43	191	18.6	915	2	O6UWA5	O6uwa5 homo sapien
44	191	18.6	937	2	Q96FT5	Q96ft5 homo sapien
45	191	18.6	956	1	MTN2_HUMAN	O00339 homo sapien

#### ALIGNMENTS

```
DR HSP; P18614; IMHP.
DR Genew; HGNC:6135; ITGA10.
DR MIW; 604042; -.
DR GO; GO:0008305; C:integrin complex; TAS.
DR GO; GO:0005518; P:collagen binding; TAS.
DR GO; GO:0007160; P:cell-matrix adhesion; TAS.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
DR PROSITE; PS0234; VWA; 1.
DR Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;
KW Repeat; Signal; Transmembrane.
FT SIGNAL 1 22 Potential.
FT CHAIN 23 1167 Integrin alpha-10.
FT DOMAIN 23 1122 Extracellular (Potential).
FT TRANSMEM 1123 1145 Potential.
FT DOMAIN 1146 1167 Cytoplasmic (Potential).
FT REPEAT 38 97 FG-GAP 1.
FT REPEAT 98 ? FG-GAP 2.
FT DOMAIN 167 350 VWF_A.
FT REPEAT 365 427 FG-GAP 3.
FT REPEAT 428 482 FG-GAP 4.
FT REPEAT 483 545 FG-GAP 5.
FT REPEAT 546 605 FG-GAP 6.
FT REPEAT 608 660 FG-GAP 7.
FT DOMAIN 1134 1140 Poly-Leu.
FT CA_BIND 494 502 Potential.
FT CA_BIND 558 566 Potential.
FT CA_BIND 620 628 Potential.
FT DISULFID 76 86 By similarity.
FT DISULFID 666 675 By similarity.
FT DISULFID 681 736 By similarity.
FT DISULFID 789 795 By similarity.
FT CARBOHYD 98 98 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 234 234 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 336 336 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 364 364 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 733 733 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 763 763 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 839 839 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 921 921 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1011 1011 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1018 1018 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 1039 1039 N-linked (GlcNAc. .) (Potential).
FT CONFLICT 844 844 I -> L (in Ref. 2).
FT CONFLICT 909 909 G -> V (in Ref. 2).
FT CONFLICT 926 926 E -> D (in Ref. 2).
SQ SEQUENCE 1167 AA; 127573 MW; AE7D3A1C25C1AE0 CRC64;

Query Match 100.0%; Score 1025; DB 1; Length 1167;
Best Local Similarity 100.0%; Pred. No. 8e-82;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GICARVDASFPQGSGLAPTAQCPTMYDVIIVLDGNSIYPWSEVQTFLRLVGLKFLIDP 60
Db 140 GICARVDASFPQGSGLAPTAQCPTMYDVIIVLDGNSIYPWSEVQTFLRLVGLKFLIDP 199

Qy 61 EOIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRREGRTKTAQAINVACTEGFSQ 120
Db 200 EOIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRREGRTKTAQAINVACTEGFSQ 259

Qy 121 SHGGRPEARLLVVVTDGESHGDELPAALKACAGRVTRYGVIAVLGHLRQRDPSSFL 180
Db 260 SHGGRPEARLLVVVTDGESHGDELPAALKACAGRVTRYGVIAVLGHLRQRDPSSFL 319

Qy 181 REIRTIASDPDERFFNV 198
Db 181 REIRTIASDPDERFFNV 198

Query Match 100.0%; Score 547; DB 2; Length 1188;
Best Local Similarity 53.0%; Pred. No. 1.9e-39;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

Qy 1 GICARVDASFPQGSGLAPTAQCPTMYDVIIVLDGNSIYPWSEVQTFLRLVGLKFLIDP 60
Db 137 GMCVRVSNFRPSKTVAPALQRCQTYMDIVILVDGNSIYPWSEVQHLINILKFKYIGP 196

Qy 61 EOIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRREGRTKTAQAINVACTEGFSQ 120
Db 197 GOIQVGLVQYGEDAVDAHEFLHNDYRSVKDVVEAAASHIEQRCGTETRTAFGEFARSEAFQ 256

Qy 121 SHGGRPEARLLVVVTDGESHGDELPAALKACAGRVTRYGVIAVLGHLRQRDPSSFL 180
Db 257 --GGRKCAKVMIVITDGESHSDPDLEKVRQSEKDNVTRYAVAVLGYNNGINPTFL 314

Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 3
ITAH HUMAN
ID ITAH HUMAN STANDARD; PRT; 1189 AA.
AC Q9UKX5; Q9UKQ1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DB Integrin alpha-11 precursor.
GN Name=ITG11;
```

OS Homo sapiens (Human),  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Fetal heart, and Osteoblast;  
 RC MEDLINE=99417678; PubMed=10486209; DOI=10.1006/geno.1999.5909;  
 RX Lehnert K., Ni J., Leung E., Gough S.M., Weaver A., Yao W.P., Liu D.,  
 RA Wang S.-X., Morris C.M., Krissansen G.W.;  
 RA "Cloning, sequence analysis, and chromosomal localization of the novel  
 RT human integrin alpha11 subunit (ITGA11).";  
 RL Genomics 60:179-187(1999).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Fetal muscle, and Uterus;  
 RC MEDLINE=99395147; PubMed=10464311; DOI=10.1074/jbc.274.36.25735;  
 RX Velling T., Kusche-Gullberg M., Seijersen T., Gullberg D.;  
 RA "cDNA Cloning and Chromosomal Localization of Human alpha(11)  
 RT Integrin. A collagen-binding, i domain-containing, beta(1)-associated  
 RL integrin alpha-chain present in muscle tissues.";  
 RL J. Biol. Chem. 274:25735-25742(1999).  
 RN [3]  
 RN SEQUENCE OF 954-1188 FROM N.A.  
 RP TISSUE=Fibroblast;  
 RC Andreu N., Estivill X., Escarceller M., Sunoy L.;  
 RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Integrin alpha-11/beta-1 is a receptor for collagen.  
 CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-11  
 CC associates with beta-1.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- TISSUE SPECIFICITY: According Ref.2 highest levels in uterus and  
 CC heart, intermediate levels in skeletal muscle and intermediate to  
 CC low levels in pancreas, kidney and placenta. According to Ref.1  
 CC also found in brain, colon, lung, small intestine, stomach,  
 CC testis, salivary glands, thyroid glands and prostate. Very low  
 CC levels in peripheral blood lymphocytes, fetal brain and fetal  
 CC liver.  
 CC -!- DEVELOPMENTAL STAGE: Strongly up-regulated in differentiating  
 CC fetal muscle cells (in vitro).  
 CC -!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins  
 CC with I-domains do not undergo protease cleavage.  
 CC -!- SIMILARITY: Belongs to the integrin alpha chain family.  
 CC -!- SIMILARITY: Contains 7 FG-GAP repeats.  
 CC -!- SIMILARITY: Contains 1 VWFA domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF109681; AA01258.1; -;  
 CC EMBL; AF137378; AAD51919.2; -;  
 CC EMBL; AL359064; CAB94392.1; -;  
 CC HSSP; P18614; 1MHP.  
 CC Genew; HGNC:6136; ITGA11.  
 CC MW; 604789; -;  
 CC GO; GO:0008305; C:integrin complex; TAS.  
 CC GO; GO:0005518; F:collagen binding; TAS.  
 CC GO; GO:0007160; P:cell-matrix adhesion; TAS.  
 CC GO; GO:0007517; P:muscle development; TAS.  
 CC InterPro; IPR000413; Integrin\_alpha.  
 CC InterPro; IPR002035; VWFA.  
 CC Pfam; PF01839; FG-GAP; 3.  
 CC Pfam; PF00092; VWA; 1.  
 CC PRINTS; PR01185; INTEGRINA.  
 CC PRINTS; PR00453; VWFADOMAIN.  
 CC SMART; SM00191; Int\_alpha; 5.  
 CC SMART; SM00327; VWA\_1.  
 CC PROSITE; PS00242; INTEGRIN\_ALPHA; FALSE\_NEG.

DR PROSITE; PS50234; VWFA; 1.  
 KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium;  
 KW Polymorphism; Receptor; Repeat; Signal; Transmembrane.  
 FT SIGNAL 1 22 Potential.  
 FT CHAIN 23 1189 Integrin alpha-11.  
 FT DOMAIN 23 1142 Extracellular (Potential).  
 FT TRANSMEM 1143 1165 Potential.  
 FT DOMAIN 1166 1189 Cytoplasmic (Potential).  
 FT REPEAT 38 94 FG-GAP 1.  
 FT DOMAIN 102 163 FG-GAP 2.  
 FT REPEAT 164 345 VWFA.  
 FT DOMAIN 359 420 FG-GAP 3.  
 FT REPEAT 422 475 FG-GAP 4.  
 FT REPEAT 477 537 FG-GAP 5.  
 FT REPEAT 539 598 FG-GAP 6.  
 FT DOMAIN 601 653 FG-GAP 7.  
 FT DOMAIN 1154 1177 Poly-Leu.  
 FT DOMAIN 1174 1177 Poly-Arg.  
 FT CA\_BIND 488 496 Potential.  
 FT CA\_BIND 551 559 Potential.  
 FT CA\_BIND 613 621 Potential.  
 FT DISULFID 76 83 By similarity.  
 FT DISULFID 121 139 Potential.  
 FT DISULFID 129 159 Potential.  
 FT DISULFID 659 668 By similarity.  
 FT DISULFID 674 729 By similarity.  
 FT DISULFID 781 787 By similarity.  
 FT DISULFID 881 893 By similarity.  
 FT CARBOHYD 82 82 N-linked (GlcNAc..)  
 FT CARBOHYD 95 95 N-linked (GlcNAc..)  
 FT CARBOHYD 291 291 N-linked (GlcNAc..)  
 FT CARBOHYD 331 331 N-linked (GlcNAc..)  
 FT CARBOHYD 358 358 N-linked (GlcNAc..)  
 FT CARBOHYD 449 449 N-linked (GlcNAc..)  
 FT CARBOHYD 462 462 N-linked (GlcNAc..)  
 FT CARBOHYD 528 528 N-linked (GlcNAc..)  
 FT CARBOHYD 642 642 N-linked (GlcNAc..)  
 FT CARBOHYD 694 694 N-linked (GlcNAc..)  
 FT CARBOHYD 857 857 N-linked (GlcNAc..)  
 FT CARBOHYD 894 894 N-linked (GlcNAc..)  
 FT CARBOHYD 973 973 N-linked (GlcNAc..)  
 FT CARBOHYD 1032 1032 N-linked (GlcNAc..)  
 FT CARBOHYD 1040 1040 N-linked (GlcNAc..)  
 FT VARIANT 433 433 V->M.  
 FT VARIANT 524 524 /FTID=VAR\_009889.  
 FT VARIANT 972 972 /FTID=VAR\_009890.  
 FT VARIANT 1003 1003 /FTID=VAR\_009891.  
 FT VARIANT 1030 1030 /FTID=VAR\_009892.  
 FT VARIANT 1094 1094 Missing.  
 FT VARIANT 1094 1094 /FTID=VAR\_009893.  
 FT VARIANT 1094 1094 L->V.  
 FT VARIANT 1094 1094 /FTID=VAR\_009894.  
 FT SEQUENCE 1189 AA; 133609 MW; 60303C08A44CD52 CRC64;  
 Query Match 53.1%; Score 544; DB 1; Length 1189;  
 Best Local Similarity 53.0%; Pred. No. 3.5e-39;  
 Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;  
 Qy 1 GICARVDASFPQGSGLAPTRQCTYMDVIVLDGNSIYVPSVQVFLRLVGLFIDP 60  
 Db 137 GMSRVNSNFRFSKTVAPALQRCQTYMDIVLDGNSIYVPSVQVFLRLVGLFIDP 196  
 Qy 61 EQIQVGLVQYGESPVHWSLGDFTKEEVRAAKNLSRREGTKTAAIMVACTEGFSQ 120  
 Db 197 GQIQVGVVQYGEDVDFHFLNDYKSDVVEAASIEQRGTETRTAFGIEFASEAFQK 256  
 Qy 121 SHGGRPEAARLLVVVTDGESHGDSGELPAALKACACAGRVTRYGIADVGLHYLRQRDPSSFL 180  
 Db 257 --GGRKGAKKMVIITDGESHSDSPDLKVKVQCSERDNDVTRVAVAVLGYNNGINPETFL 314



```

Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 4
ID ITAH MOUSE STANDARD; PRT; 1188 AA.
AC P61622;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DE Integrin alpha-11 precursor.
GN Name=ITGAL;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haie H.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci F., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wexley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fehey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Brakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Kravinsky M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Integrin alpha-11/beta-1 is a receptor for collagen (By
CC similarity).
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-11
CC associates with beta-1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; BC058716; AAHS8716.1; -.
CC MGD; MGI:2442114; ITGAL1.
CC PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
CC PROSITE; PS02234; VWFA; 1.
CC Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;
CC Repeat; Signal; Transmembrane.
CC SIGNAL 1 22 Potential.
CC CHAIN 23 1188 Integrin alpha-11.
CC DOMAIN 23 1141 Extracellular (Potential).
CC TRANSMEM 1142 1164 Potential.
CC DOMAIN 1165 1188 Cytoplasmic (Potential).
CC REPEAT 38 94 FG-GAP 1.
FT REPEAT 102 163 FG-GAP 2.
FT DOMAIN 164 345 VWFA.
FT REPEAT 359 420 FG-GAP 3.
FT REPEAT 422 475 FG-GAP 4.
FT REPEAT 477 537 FG-GAP 5.
FT REPEAT 539 598 FG-GAP 6.
FT REPEAT 601 653 FG-GAP 7.
FT CA_BIND 551 559 Potential.
FT CA_BIND 613 621 Potential.
FT DISULFID 76 83 By similarity.
FT DISULFID 121 139 Potential.
FT DISULFID 129 159 Potential.
FT DISULFID 659 668 By similarity.
FT DISULFID 674 729 By similarity.
FT DISULFID 781 787 By similarity.
FT DISULFID 881 893 By similarity.
FT CARBOHYD 82 882 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 95 95 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 291 291 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 331 331 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 358 358 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 449 449 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 462 462 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 528 528 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 642 642 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 694 694 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 857 857 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 894 894 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 973 973 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 1031 1031 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 1039 1039 N-linked (GLNAC. .) (Potential).
FT CARBOHYD 1059 1059 N-linked (GLNAC. .) (Potential).
SQ SEQUENCE 1188 AA; 133011 MW; 080313C90D65422E CRC64;

Query Match 53.0%; Score 543; DB 1; Length 1188;
Best Local Similarity 52.5%; Pred. No. 4.2e-39;
Matches 104; Conservative 39; Mismatches 53; Indels 2; Gaps 1;

Qy 1 GICARVDASFOGSLAPTAQRCPTVMDVIVLDGNSIYPWSEVQTFRLRLVGLKFLDP 60
Db 137 GMCSTRANSFRFSKTVAPALQRCQTYMDIVIVLDGNSIYPWSEVQHFLLNLKKFYGP 196

Qy 61 EQQVGLVQVQGESPVHWSLGDPRTEKEVVVRAAKNLRSRREGRTKTAQALMVACTEGFSQ 120
Db 197 GQIQVGIQVQGEDAVHEFHLNDYRSVKDVVEASHIEQRCGTETRTAFGIEFARSEAFQK 256

Qy 121 SHGGRPEAARLLVVVTDGSHDGEELPAALKACEAGRVTRYGIYGLHYLRQRDPSSFL 180
Db 257 --GGRKGAKKVMIVITDGHSDSPDLKQVIRQSEKDNVTRYAVAVLVGYNNRRGINPETFL 314

Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 5
ID ITAH_CHICK STANDARD; PRT; 285 AA.
AC Q90615;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a)
DE (Fragment).
GN Name=ITGAL;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.

```

RC TISSUE=Gizzard;  
 RX MEDLINE=94357930; PubMed=7521332;  
 RA Kern A., Briesewitz R., Bank I., Marcantonio E.B.;  
 RT "The role of the I domain in ligand binding of the human integrin  
 alpha 1 beta 1";  
 RL J. Biol. Chem. 269:22811-22816(1994).  
 CC -I- FUNCTION: Integrin alpha-1/beta-1 is a receptor for laminin and  
 collagen. It recognizes the proline-hydroxylated sequence G-F-P-G-  
 E-R in collagen.  
 CC -I- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-1  
 associates with beta-1.  
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -I- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins  
 with I-domains do not undergo protease cleavage.  
 CC -I- SIMILARITY: Belongs to the integrin alpha chain family.  
 CC -I- SIMILARITY: Contains 1 VWFA domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U10114; AAA59067.1; --  
 DR HSP; P18614; IMHP.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00092; VWA; 1.  
 DR PRINTS; PR00453; VMFADOMAIN.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; PARTIAL.  
 DR PROSITE; PS0234; VWFA; 1.  
 KW Cell adhesion; Glycoprotein; Integrin; Receptor; Transmembrane.  
 FT NON\_TER 1  
 FT DOMAIN 1  
 FT DOMAIN <1>285 Extracellular (Potential).  
 FT CARBOHYD 66 279 VWFA.  
 FT CARBOHYD 2 2 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 40 40 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 208 208 N-linked (GlcNAc. . .) (Potential).  
 FT CARBOHYD 232 232 N-linked (GlcNAc. . .) (Potential).  
 FT NON\_TER 285 285  
 SQ SEQUENCE 285 AA; 31503 MW; 1B05D3246CD5CA7E CRC64;  
 Query Match 46.1%; Score 473; DB 1; Length 285;  
 Best Local Similarity 45.5%; Pred. No. 1.2e-33;  
 Matches 90; Conservative 37; Mismatches 71; Indels 0; Gaps 0;  
 Qy 1 GICARVDASFOQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKLFDP 60  
 Db 36 GVCNSVSTFTETVKAVAPSVQECKTQLDIVIVLDGNSIYPWSEVQTFLLRLVGLKLFDP 95  
 Qy 61 EQIQVGLVQYGESPVHEWSLGDFTKEEVVRAAKNLSRREGRETAKQAQIMVACTEGFSQ 120  
 Db 96 QQTQVGIVQYGVQVWVHEFLYNTYSTTBEVMDAALRIQRGTQTMALGIDTAREEAFTE 155  
 Qy 121 SHGCRPEAARLLVVVTGESHGDEELPAALKACBAGRVTRYGIAGVGLHYLRQRDPSSFL 180  
 Db 156 AHGARRGVQKVMVIVTGDESHNDYRLQEVIDKCEDENIQRFALIGSYSGNLSSTKFEV 215  
 Qy 181 REINTIASDPPERFFNV 198  
 Db 216 EEIKSIASKPTEKHFFNV 233  
 RESULT 6  
 O42094  
 ID O42094 PRELIMINARY; PRT; 1171 AA.  
 AC O42094;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JAN-2004 (TrEMBLrel. 26, Last annotation update)

DE Alpha integrin.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Gizzard;  
 RX MEDLINE=97476270; PubMed=9334246; DOI=10.1074/jbc.272.42.26643;  
 RA Obata H., Hayaashi K., Nishida W., Momiyama T., Uchida A., Ochi T.,  
 RA Sobue K.;  
 RT "Smooth muscle cell phenotype-dependent transcriptional regulation of  
 the alpha1 integrin gene";  
 RL J. Biol. Chem. 272:26643-26651(1997).  
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -I- SIMILARITY: Belongs to the integrin alpha chain family.  
 DR EMBL; AB000470; BAA23160.1; --  
 DR PIR; A55348; A55348.  
 DR HSP; P18614; IMHP.  
 DR GO; GO:0008305; C:integrin complex; IEA.  
 DR GO; GO:0007160; P:cell-matrix adhesion; IEA.  
 DR GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 1.  
 DR Pfam; PF00357; Integrin\_alpha; 1.  
 DR Pfam; PF00092; VWA; 1.  
 DR PRINTS; PR01185; INTEGRINA.  
 DR PRINTS; PR00453; VMFADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 4.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS0234; VWFA; 1.  
 KW Cell adhesion; Integrin; Transmembrane.  
 SQ SEQUENCE 1171 AA; 130228 MW; B505A4D65F09736E CRC64;  
 Query Match 46.1%; Score 473; DB 2; Length 1171;  
 Best Local Similarity 45.5%; Pred. No. 6.7e-33;  
 Matches 90; Conservative 37; Mismatches 71; Indels 0; Gaps 0;  
 Qy 1 GICARVDASFOQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKLFDP 60  
 Db 135 GVCNSVSTFTETVKAVAPSVQECKTQLDIVIVLDGNSIYPWSEVQTFLLRLVGLKLFDP 194  
 Qy 61 EQIQVGLVQYGESPVHEWSLGDFTKEEVVRAAKNLSRREGRETAKQAQIMVACTEGFSQ 120  
 Db 195 QQTQVGIVQYGVQVWVHEFLYNTYSTTBEVMDAALRIQRGTQTMALGIDTAREEAFTE 254  
 Qy 121 SHGCRPEAARLLVVVTGESHGDEELPAALKACBAGRVTRYGIAGVGLHYLRQRDPSSFL 180  
 Db 255 AHGARRGVQKVMVIVTGDESHNDYRLQEVIDKCEDENIQRFALIGSYSGNLSSTKFEV 314  
 Qy 181 REINTIASDPPERFFNV 198  
 Db 315 EEIKSIASKPTEKHFFNV 332  
 RESULT 7  
 ITAL\_HUMAN  
 ID ITAL\_HUMAN STANDARD; PRT; 1151 AA.  
 AC P56199;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a).  
 GN Name=ITGAL;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=93155124; PubMed=8428973;  
 RA Briesewitz R., Epstein M.R., Marcantonio E.E.;  
 RT "Expression of native and truncated forms of the human integrin alpha  
 1 subunit.";  
 RL J. Biol. Chem. 268:2989-2996(1993).  
 CC -!- FUNCTION: Integrin alpha-1/beta-1 is a receptor for laminin and  
 CC collagen. It recognizes the proline-hydroxylated sequence G-P-P-G-  
 CC E-R in collagen.  
 CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-1  
 CC associates with beta-1.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins  
 CC with I-domains do not undergo protease cleavage.  
 CC -!- SIMILARITY: Belongs to the integrin alpha chain family.  
 CC -!- SIMILARITY: Contains 7 FG-GAP repeats.  
 CC -!- SIMILARITY: Contains 1 VWFA domain.  
 CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD49a entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd49a.htm".  
 DR PTR; A45226; A45226.  
 DR PDB; 1QC5; X-ray; A/B=140-331.  
 DR PDB; 1QCY; X-ray; A=141-333.  
 DR Genew; HGNC:6134; ITGAL.  
 DR MIM; 192868;  
 DR GO; GO:0008305; C:integrin complex; TAS.  
 DR GO; GO:0005518; F:collagen binding; TAS.  
 DR GO; GO:0007160; P:cell-matrix adhesion; NAS.  
 DR InterPro; IPR004113; Integrin\_alpha.  
 DR InterPro; IPR002035; VWFA.  
 DR Pfam; PF01839; FG-GAP; 3.  
 DR Pfam; PF00357; Integrin\_alpha; 1.  
 DR Pfam; PF00092; VWA; 1.  
 DR PRINTS; PR01185; INTEGRINA.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 5.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS0234; VWFA; 1.  
 KW 3D-structure; Calcium; Cell adhesion; Glycoprotein; Integrin;  
 KW Magnesium; Receptor; Repeat; Transmembrane.  
 FT DOMAIN 1 1113 Extracellular (Potential).  
 FT TRANSMEM 1114 1136 Potential.  
 FT DOMAIN 1137 1151 Cytoplasmic (Potential).  
 FT REPEAT 16 75 FG-GAP 1.  
 FT REPEAT 76 ? FG-GAP 2.  
 FT REPEAT 147 360 VWFA.  
 FT REPEAT 349 404 FG-GAP 3.  
 FT REPEAT 405 457 FG-GAP 4.  
 FT REPEAT 459 520 FG-GAP 5.  
 FT REPEAT 540 599 FG-GAP 6.  
 FT REPEAT 602 654 FG-GAP 7.  
 FT CA\_BIND 470 478 Potential.  
 FT CA\_BIND 552 560 Potential.  
 FT CA\_BIND 614 622 Potential.  
 FT SITE 1139 1142 GFFKR motif.  
 FT DISULFID 54 64 By similarity.  
 FT DISULFID 660 669 By similarity.  
 FT DISULFID 675 728 By similarity.  
 FT DISULFID 780 786 By similarity.  
 FT DISULFID 850 858 By similarity.  
 FT DISULFID 1002 1034 By similarity.  
 FT DISULFID 1037 1044 By similarity.  
 FT CARBOHYD 46 46 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 72 72 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 77 77 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 84 84 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 189 189 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 289 289 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 313 313 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 374 374 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 390 390 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 432 432 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 504 504 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 671 671 N-linked (GlcNAc... ) (Potential).

FT CARBOHYD 720 720 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 752 752 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 812 812 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 855 855 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 880 880 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 887 887 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 911 911 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 938 938 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 946 946 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 980 980 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1045 1045 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1055 1055 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1074 1074 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1085 1085 N-linked (GlcNAc... ) (Potential).  
 FT STRAND 143 150  
 FT TURN 153 154  
 FT TURN 158 158  
 FT TURN 159 169  
 FT TURN 171 171  
 FT STRAND 174 174  
 FT TURN 176 177  
 FT STRAND 179 186  
 FT STRAND 190 194  
 FT TURN 196 197  
 FT TURN 202 211  
 FT HELIX 222 231  
 FT TURN 232 234  
 FT HELIX 236 238  
 FT TURN 239 239  
 FT TURN 242 243  
 FT STRAND 245 252  
 FT TURN 257 258  
 FT HELIX 259 261  
 FT HELIX 262 271  
 FT TURN 272 273  
 FT STRAND 274 281  
 FT HELIX 283 287  
 FT TURN 288 289  
 FT HELIX 293 302  
 FT HELIX 307 310  
 FT STRAND 311 314  
 FT HELIX 317 323  
 FT HELIX 324 330  
 SQ SEQUENCE 1151 AA; 127837 MW; 6B3F3C1AABF52808 CRC64;  
 Query Match 45.6%; Score 467.5; DB 1; Length 1151;  
 Best Local Similarity 46.5%; Pred. No. 2e-32;  
 Matches 92; Conservative 36; Mismatches 69; Indels 1; Gaps 1;  
 Qy 1 GICARVDASFPQGS LAPTAQRCPYMDVVIVLDGSGNSIYPWSEVQTFLLRLVGLKLFIDP 60  
 Db 118 GICSDVSPTQVNVSIAP-VQECSTQLDIVLDGSGNSIYPWSDVTAFLNDLLKRWIDGP 176  
 Qy 61 EQIQVGLVQYGESPVHWSLGDFTKBEVVVRAAKNLSRRREGRTKTAQAIWVACTEFSQ 120  
 Db 177 KQIQVGIQYQGVNTHFNLKYSSTEEVLVAAKKIVQGRGRTMTALGDTARKEAFTE 236  
 Qy 121 SHGGRPEARLLVVVTDGESHGDBELPAALKACEAGRVTRYGIATLGHVLRQRDPSSFL 180  
 Db 237 ARGARRGVKKVMVITVDGESHNRHLKVKVQDCEDENIQFISAILGSGYNRGNLSKTFV 296  
 Qy 181 REIRTIASDDPDERFFNV 198  
 Db 297 EEIKSIASEPTEKHFFNV 314  
 RESULT 8  
 ID ITAI\_RAT STANDARD; PRT; 1180 AA.  
 AC P19614;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)



```

Qy 61 EQIQVGLVQGESPVHWSLGDFTKBEVRAAKNLSRRREGRETAKQAIAWVACTGFSQ 120
Db 205 KQTQGVQVQGENVTHEFNKNKYSSTBEVLVAANKIGRQGLQMTALGIDTARKEAFTE 264
Qy 121 SHGGRPBAARLLVVVTGESHGDBELPAALAKACBAGRVTRYGIAVLGHYLRORRQDPSFL 180
Db 265 ARGARRGVKKVMVVTGESHGDBELPAALAKACBAGRVTRYGIAVLGHYLRORRQDPSFL 324
Qy 181 REIRTIASDPDERPFNNV 198
Db 325 BEIKSIASEPTEKHFNNV 342

RESULT 9
ITA2 MOUSE STANDARD; PRT; 1178 AA.
AC Q62469; Q62163;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GPIa)
DE (Collagen receptor) (VLA-2 alpha chain) (CD49b).
GN Name=Itga2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX MEDLINE=94363406; PubMed=8081889;
RA Edelman J.M., Chan B.M., Unival S., Onodera H., Wang D.Z.,
RA Danjanovich L., Latzer D.B., Finberg R.W., Bergelson J.M.;
RT "The mouse VLA-2 homologue supports collagen and laminin adhesion but
RL not virus binding.";
RN [2]
RP SEQUENCE OF 450-1178 FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=94355691; PubMed=7521231;
RA Wu J.B., Santoro S.A.;
RT "Complex patterns of expression suggest extensive roles for the alpha
RL 2 beta 1 integrin in murine development.";
CC Dev. Dyn. 199; 292-314 (1994).
CC 1- FUNCTION: Integrin alpha-2/beta-1 is a collagen receptor, being
CC responsible for adhesion of platelets and other cells to
CC collagen, modulation of collagen and collagenase gene expression,
CC force generation and organization of newly synthesized
CC extracellular matrix. It is also a receptor for laminins, collagen
CC C-propeptides and E-cadherin. Mice homozygous for a null mutation
CC in the alpha-2 die very early in embryogenesis.
CC 1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-2
CC associates with beta-1.
CC 1- SUBCELLULAR LOCATION: Type I membrane protein.
CC 1- DOMAIN: The integrin I-domain (insert) is a VMPA domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC 1- SIMILARITY: Belongs to the integrin alpha chain family.
CC 1- SIMILARITY: Contains 7 FG-GAP repeats.
CC 1- SIMILARITY: Contains 1 VMPA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z29987; CAA82877.1; -;
CC EMBL; X75427; CAA53178.1; -;
CC FIC; S44142; S44142.
CC HSSP; P17301; LAOX.

```

```

DR MGD; MGI:96600; Itga2.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 2.
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWF_A; 1.
KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Platelet;
KW Receptor; Repeat; Signal; Transmembrane.
FT SIGNAL 1 26 By similarity.
FT CHAIN 27 1178 Integrin alpha-2.
FT DOMAIN 27 1129 Extracellular (Potential).
FT TRANSMEM 1130 1151 Potential.
FT DOMAIN 1152 1178 Cytoplasmic (Potential).
FT REPEAT 42 100 FG-GAP 1.
FT REPEAT 101 7 FG-GAP 2.
FT DOMAIN 185 375 VMPA.
FT REPEAT 376 430 FG-GAP 3.
FT REPEAT 431 483 FG-GAP 4.
FT REPEAT 485 546 FG-GAP 5.
FT REPEAT 548 607 FG-GAP 6.
FT REPEAT 612 664 FG-GAP 7.
FT CA_BIND 496 504 Potential.
FT CA_BIND 560 568 Potential.
FT CA_BIND 624 632 Potential.
FT SITE 480 482 Cell attachment site (Potential).
FT SITE 1154 1158 GPPK motif.
FT DISULFID 80 89 By similarity.
FT DISULFID 677 734 By similarity.
FT DISULFID 786 792 By similarity.
FT DISULFID 862 873 By similarity.
FT DISULFID 1016 1047 By similarity.
FT DISULFID 1052 1057 By similarity.
FT CARBOHYD 102 102 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 109 109 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 429 429 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 457 457 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 472 472 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 696 696 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1054 1054 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1071 1071 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1078 1078 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 1178 AA; 128926 MW; 1F194B9C0240F465 CRC64;

Query Match 44.3%; Score 454; DB 1; Length 1178;
Best Local Similarity 46.0%; Pred. No. 3-2e-31;
Matches 91; Conservative 33; Mismatches 74; Indels 0; Gaps 0;

Qy 1 GICARVDASFPQGSAPTAQRCTYMDVVIVLDGNSIYPMSEVQTFRLRLVGLKFLDP 60
Db 144 GICSDVSPDFQFLTSFPAQVACPSLVDVVVVCDSNSIYPMSEVAVKFLVKFTGLDIP 203
Qy 61 EQIQVGLVQGESPVHWSLGDFTKBEVRAAKNLSRRREGRETAKQAIAWVACTGFSQ 120
Db 204 KKTQVALIQVANEPRIFINLDFETKEDMVOATSETRQHGCDLNTFRLEFARDVAYSQ 263
Qy 121 SHGGRPBAARLLVVVTGESHGDBELPAALAKACBAGRVTRYGIAVLGHYLRORRQDPSFL 180
Db 264 TSGRPGATKVMVVTGESHGDBELPAALAKACBAGRVTRYGIAVLGHYLRORRQDPSFL 323
Qy 181 REIRTIASDPDERPFNNV 198
Db 324 KEIKAIASPTETRYFNNV 341

RESULT 10
Q6P1C7 PRELIMINARY; PRT; 1178 AA.
ID Q6P1C7

```

```

AC 06P1C7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Integrin alpha 2.
GN Name=Itga2;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Eye;
RX MEDLINE=223825; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.F., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.M., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchkov L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Woxley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Eye;
RA Strausberg R.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
DR EMBL; BC065113; AAH51139.1; --
DR HSSP; P17301; IAOX.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0016021; C:integral to membrane; TAS.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 2_
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA_1
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWF; 1.
DR Cell adhesion; Integrin; Transmembrane.
KW Cell adhesion; Integrin; Transmembrane.
SQ SEQUENCE 1178 AA; 128954 MW; 62FAEA820242A9B6 CRC64;

Query Match 44.3%; Score 454; DB 2; Length 1178;
Best Local Similarity 46.0%; Pred. No. 3.2e-31;
Matches 91; Conservative 33; Mismatches 74; Indels 0; Gaps 0;

Qy 1 GICARVDASFOQGSAPTAQRCPTMDVVLVLDGNSIYPWSEVQTFELRLVGLKFLDIP 60
Dy 144 GICSDVDFQFLTSFSPAQACSLVDVVVCDENSIYPWEAVKFLVKFVLGDLGIP 203
Qy 61 EIQVGLVQGESPVHESLGDFTKEEVRAAKNLSRREGRTKTAQIMVACTGFSQ 120
Dy 204 KKTQVALIQYANEPRIIFENLDFETKEDMWQATSETRQHGGLDNTFRAIEFARDYAYSQ 263
Qy 121 SHGGRPEARLLVVVTDGSHGDLGELPAALKACEAGRTYRGIAVLGHYLRDRDPSSFL 180

```

```

Db 264 TSGRPGATKVMVVVTDGSHGDKLKVITQCCNDDEILRFGIAVLGYLNRNALDTKNLI 323
Qy 181 REINTIASDPDERFFNV 198
Dy 324 KEIKAIASPTPTERYFFNV 341

RESULT 11
ID ITA2_BOVIN STANDARD; PRT; 1170 AA.
AC P53710;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GP1a)
DE (Collagen receptor) (VLA-2 alpha chain) (CD49b) (Fragment).
GN Name=ITGA2;
OS Bos taurus (Bovine);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94193647; PubMed=7511592;
RA Kamata T., Puzon W., Takada Y.;
RT "Identification of putative ligand binding sites within I domain of
RT integrin alpha 2 beta 1 (VLA-2, CD49b/CD29).";
RL J. Biol. Chem. 269:9659-9663(1994).
CC -1- FUNCTION: Integrin alpha-2/beta-1 is a receptor for laminin.
CC collagen, collagen C-propeptides, fibronectin and E-cadherin. It
CC recognizes the proline-hydroxylated sequence G-P-G-E-R in
CC collagen. It is responsible for adhesion of platelets and other
CC cells to collagens, modulation of collagen and collagenase gene
CC expression, force generation and organization of newly synthesized
CC extracellular matrix.
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-2
CC associates with beta-1.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DOMAIN: The integrin I-domain (insert) is a VWF domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
CC -1- SIMILARITY: Contains 1 VWF domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; L25886; AAB59255.1; --
DR PIR; I45914; I45914.
DR HSSP; P17301; IAOX.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 2_
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA_1
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; VWF; 1.
KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Platelet;
KW Polymorphism; Receptor; Repeat; Signal; Transmembrane.
FT NON_TER 1 1
FT SIGNAL <1 18 Potential.
FT CHAIN 19 1170 Extracellular (Potential).
FT DOMAIN 19 1121

```





DR PDB; 1DZ1; X-ray; A=171-355.  
 DR Genew; HGNC:6137; ITGA2.  
 DR MM; 192374; .  
 DR GO; GO:0008305; C:integrin complex; TAS.  
 DR GO; GO:0005886; C:plasma membrane; TAS.  
 DR GO; GO:0005518; F:collagen binding; TAS.  
 DR GO; GO:0007596; P:blood coagulation; TAS.  
 DR GO; GO:0007160; P:cell-matrix adhesion; TAS.  
 DR GO; GO:0009887; P:organogenesis; TAS.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR02035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 2.  
 DR Pfam; PF00357; Integrin\_alpha; 1.  
 DR Pfam; PF00092; VWA; 1.  
 DR PRINTS; PRO1185; INTEGRIN.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 5.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS0234; VFWA; 1.  
 DR 3D-structure; Calcium; Cell adhesion; Direct protein sequencing;  
 KW Glycoprotein; Integrin; Magnesium; Platelet; Polymorphism; Receptor;  
 KW Repeat; Signal; Transmembrane.  
 FT SIGNAL 1 29  
 FT CHAIN 30 1181 Integrin alpha-2.  
 FT DOMAIN 1133 1154 Extracellular (Potential).  
 FT TRANSMEM 1133 1154 Potential.  
 FT DOMAIN 1155 1181 Cytoplasmic (Potential).  
 FT DOMAIN 1155 1161 Interaction with HP85.  
 FT REPEAT 45 103 FG-GAP 1.  
 FT REPEAT 104 7 FG-GAP 2.  
 FT REPEAT 188 378 VFWA.  
 FT REPEAT 379 433 FG-GAP 3.  
 FT REPEAT 434 486 FG-GAP 4.  
 FT REPEAT 488 549 FG-GAP 5.  
 FT REPEAT 551 610 FG-GAP 6.  
 FT REPEAT 615 667 FG-GAP 7.  
 FT CA\_BIND 499 507 Potential.  
 FT CA\_BIND 563 571 Potential.  
 FT CA\_BIND 627 635 Potential.  
 FT SITE 1157 1161 GFPR motif.  
 FT DISULFID 83 92 By similarity.  
 FT DISULFID 680 737 By similarity.  
 FT DISULFID 789 795 By similarity.  
 FT DISULFID 865 876 By similarity.  
 FT DISULFID 1019 1050 By similarity.  
 FT DISULFID 1055 1060 By similarity.  
 FT CARBOHYD 105 105 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 112 112 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 343 343 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 432 432 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 460 460 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 475 475 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 699 699 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1057 1057 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1074 1074 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1081 1081 N-linked (GlcNAc... ) (Potential).  
 FT VARIANT 534 534 K -> E (in alloantigen HPA-5B;  
 /FTid=VAR\_003977.  
 FT TURN 170 171  
 FT STRAND 173 180  
 FT TURN 183 184  
 FT HELIX 188 199  
 FT TURN 200 201  
 FT STRAND 204 204  
 FT TURN 206 207  
 FT STRAND 209 216  
 FT STRAND 220 224  
 FT TURN 226 228  
 FT HELIX 232 240  
 FT TURN 241 241  
 FT HELIX 252 262

FT TURN 263 264  
 FT HELIX 266 268  
 FT TURN 269 269  
 FT STRAND 275 282  
 FT HELIX 289 291  
 FT HELIX 292 301  
 FT TURN 302 303  
 FT STRAND 304 311  
 FT HELIX 313 317  
 FT TURN 318 319  
 FT HELIX 323 330  
 FT TURN 331 332  
 FT HELIX 337 340  
 FT STRAND 341 344  
 FT HELIX 347 353  
 FT HELIX 354 362  
 FT TURN 363 363  
 SQ SEQUENCE 1181 AA; 129295 MW; 7E1B7ED968A94070 CRC64;  
 Query Match 43.0%; Score 441; DB 1; Length 1181;  
 Best local Similarity 44.4%; Pred. No. 4.6e-30;  
 Matches 88; Conservative 33; Mismatches 77; Indels 0; Gaps 0;  
 QY 1 GICARVDASFOQGSILAPTAQRCPYMDVIVLDGNSIYPWSEVQTFRLRLVGLKLFIDP 60  
 DB 147 GVCSDSIPDFQLSASFATQPCPSLIDVVVVCDESNIYPWDVAVKVFLEKVGQLDIGP 206  
 QY 61 EQIQVGLVQYGESPVHEWSLGDFTKKEVVRAAKNLSRREGRETKTAAQIMVACTEGFSQ 120  
 DB 207 TKTVQVGLIQYANNPRVVFNLTNTYKTEEMIVATVTSQYGGDLTNTFCAIQYARKAYSA 266  
 QY 121 SHGGRPEAARLLVVVTDGSHDGEELPAALKACAGVTRYGIAVLGHYLRQRDPSSFL 180  
 DB 267 ASGGRSATKVVVVVTDGSHDGSMLKAVIDQCNDHNLIRFGIAVLGYLNRNLDTKNLI 326  
 QY 181 REIRTIASDPDERFFNV 198  
 DB 327 KEIKAIASIPTRYFFNV 344  
 RESULT 13  
 ID Q8CB84 PRELIMINARY; PRT; 323 AA.  
 AC Q8CB84; 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Mus musculus adult male bone cDNA, RIKEN full-length enriched library,  
 clone:9830137A04 product:integrin alpha 2, full insert sequence.  
 DE (Fragment).  
 GN Name=Itga2;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Bone;  
 RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Bone;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Bone;  
 RA The FANTOM Consortium;



DR Pfam: PF00092; VWA; 1.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS0234; VWA; 1.  
KW Integrin.  
FT NON\_TER 1  
FT NON\_TER 156  
SQ SEQUENCE 156 AA; 17042 MW; BC7ABD63226C652B CRC64;

Query Match 29.2%; Score 299; DB 2; Length 156;  
Best Local Similarity 44.1%; Pred. No. 1.5e-18;  
Matches 63; Conservative 24; Mismatches 56; Indels 0; Gaps 0;

Qy	56	LFDIDPEIQVGLVQYGESPVHWSLGDFTKKEVVRAAKNLSRREGRETKTQAQIMVACT	115
Db	3	LDIGPTKTQVGLIQYANNPRVFNLTNTYKTEEMIVATSTQSYGGDLTNTFGAIQYARK	62
Qy	116	EGFSQSHGGRPEAARLLVWVDGSHDGEELPAALKACEAGRVTRYGIIVLGHVLRQRD	175
Db	63	YAYSAAAGGRRSATKVMVVVDGSHDGSMLKAVIDQCNDHNLIRFGIIVLGYLNRNALD	122
Qy	176	PSSFLEIRTIASDPDERFFNV	198
Db	123	TKNLIKSIAISIPTRYFFNV	145

Search completed: April 6, 2005, 12:15:27  
Job time : 31.3046 secs

100%

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 25.0407 Seconds  
(without alignments)  
3058.161 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337  
Perfect score: 1025  
Sequence: 1 GICARVDASFOGSLAPTA.....FLREIRTIASDPDERFFNV 198

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1025	100.0	1132	3 AAY32243	Aay32243 Human int
2	1025	100.0	1152	4 AAB64657	Aab64657 Human sec
3	1025	100.0	1152	4 AAB64658	Aab64658 Human sec
4	1025	100.0	1167	3 AAY32242	Aay32242 Human int
5	1025	100.0	1167	4 AAB64584	Aab64584 Human sec
6	1025	100.0	1167	6 ABP99490	Abp99490 Human sec
7	1025	100.0	1167	6 ABR00964	Abr00964 Human gen
8	1025	100.0	1167	6 ADA44026	Ada44026 Human sec
9	1025	100.0	1167	8 ADQ19290	Adq19290 Human sof
10	912	89.0	195	5 AAU76862	Aau76862 Human int
11	912	89.0	195	5 AAU76853	Aau76853 Human int
12	912	89.0	198	8 ADR23227	Adr23227 Human int
13	840.5	82.0	1049	5 ADR41424	Adr41424 Human CD-
14	797.5	77.8	303	3 AAY32282	Aay32282 Mouse int
15	547	53.4	1188	4 AAB50087	Aab50087 Murine A2
16	547	53.4	1188	5 AAU10552	Aau10552 Murine A2
17	544	53.1	707	4 AAU19663	Aau19663 Human nov
18	544	53.1	707	5 ABP47883	Abp47883 Human pol
19	544	53.1	707	7 ADC10845	Adc10845 Human ext
20	544	53.1	1034	3 AAB25590	Aab25590 Protein e
21	544	53.1	1034	6 ADA27062	Ada27062 Human nov
22	544	53.1	1034	8 ADE86592	Ade86592 Novel hum
23	544	53.1	1058	5 ADR41496	Adr41496 Human CD-
24	544	53.1	1188	4 AAB30929	Aab30929 Amino aci
25	544	53.1	1188	4 AAU14467	Aau14467 Human nov

26	544	53.1	1188	4 AAU14231	Aau14231 Human nov
27	544	53.1	1188	4 AAB50085	Aab50085 Human A25
28	544	53.1	1188	5 AAU10551	Aau10551 Human A25
29	544	53.1	1188	7 ADE09956	Ade09956 Novel pro
30	544	53.1	1189	3 AAB25582	Aab25582 ITGAl1 pr
31	544	53.1	1189	4 ABG12949	Abg12949 Novel hum
32	544	53.1	1189	6 ABR58364	Abr58364 Human NOV
33	544	53.1	1189	6 ADA27054	Ada27054 Human nov
34	544	53.1	1189	7 ADE63570	Ade63570 Human pro
35	544	53.1	1189	8 ADE86584	Ade86584 Novel hum
36	544	53.1	1189	8 ADQ19968	Adq19968 Human sof
37	541	52.8	437	7 ADM04701	Adm04701 Human pro
38	526.5	51.4	987	8 ADP29492	Adp29492 Human sec
39	524.5	51.2	1189	8 ADH80785	Adh80785 Human pol
40	492	48.0	193	5 AAU76863	Aau76863 Human int
41	492	48.0	193	5 AAU76854	Aau76854 Human int
42	492	48.0	196	8 ADR23228	Adr23228 Human int
43	492	48.0	1120	6 ABR58365	Abr58365 Human NOV
44	472.5	46.1	772	8 ABM80613	Abm80613 Tumour-as
45	472.5	46.1	1177	8 ADQ39498	Adq39498 Human myo

ALIGNMENTS

RESULT 1  
AAY32243  
ID AAY32243 standard; protein; 1132 AA.

XX AC AAY32243;  
XX  
DT 15-FEB-2000 (first entry)  
XX  
DE Human integrin subunit alpha-10 splice variant.

XX KW Integrin alpha-10; ISal0; human; trauma; rheumatoid arthritis;  
KW osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;  
KW therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker;  
KW splice variant.  
XX OS Homo sapiens.

XX PH Key Location/Qualifiers  
FT Peptide 1..22  
FT Protein /note= "signal peptide"  
FT /note= "mature protein"

XX WO9951639-A1.  
XX 14-OCT-1999.  
XX 31-MAR-1999; 99WO-SE0000544.  
XX 02-APR-1998; 98SE-00001164.  
XX 28-JAN-1999; 99SE-00000319.  
XX (ACTI-) ACTIVE BIOTECH AB.

XX Lundgren-Akerlund E;  
XX WPI; 2000-052639/04.  
XX N-PSDB; AAZ34720.

XX New isolated integrin subunit alpha-10, used as a marker or target molecule for cells during development, regeneration and pathological conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or inflammation.

XX Claim 1; Page 43-48; 90pp; English.

XX This sequence represents a splice variant of novel human chondrocyte integrin subunit alpha-10 (ISal0). It is identical to ISal0 (see

CC AAY32242) except for deletion of amino acids 975-986. The invention  
 CC relates to a recombinant or isolated integrin heterodimer comprising the  
 CC alpha10 subunit in association with subunit beta (especially beta-1). The  
 CC heterodimer, subunit alpha-10 or splice variant can be used as a marker  
 CC or target of all types of cells, e.g. of chondrocytes, osteoblasts and  
 CC fibroblasts. They can also be used: for treating pathological conditions  
 CC involving ISa10, such as damage to cartilage, trauma, rheumatoid  
 CC arthritis or osteoarthritis; for detecting the formation of cartilage  
 CC during embryonal development, physiological or therapeutic repair of  
 CC cartilage, or detecting regeneration of cartilage or chondrocytes during  
 CC transplantation of cartilage or chondrocytes; for selection and analysis  
 CC or for sorting, isolating or purification of chondrocytes and for in  
 CC vitro studies of differentiation of chondrocytes; and as a target for  
 CC anti-adhesive drugs or molecules in tendon, ligament, skeletal muscle or  
 CC other tissues where adhesion impairs the function of the tissue (all  
 CC claimed). ISa10 binding entities can be used to determine the  
 CC differentiation-state of cells during embryonic development, angiogenesis  
 CC or development of cancer, in pathological conditions such as rheumatoid  
 CC arthritis, osteoarthritis or cancer, in tissue regeneration or in  
 CC therapeutic and physiological repair of cartilage (claimed). A  
 CC vaccine comprising the integrin heterodimer or subunit alpha-10 is also  
 CC claimed. ISa10 polynucleotides, vectors, host cells and methods of  
 CC producing recombinant ISa10 are also claimed

XX Sequence 1132 AA;

Query Match 100.0%; Score 1025; DB 3; Length 1132;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-108;  
 Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYPMSVEVQTLRLVGLKFLDP 60  
 Db 140 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYPMSVEVQTLRLVGLKFLDP 199  
 Qy 61 EQIQVGLVQGESPVHWSLGDFTKKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQ 120  
 Db 200 EQIQVGLVQGESPVHWSLGDFTKKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQ 259  
 Qy 121 SHGGRPEAARLLVVVTDGESHGDELPALAKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180  
 Db 260 SHGGRPEAARLLVVVTDGESHGDELPALAKACEAGRVTRYGIAVLGHYLRQRDPSSFL 319  
 Qy 181 REIRTIASDPDERFFNNV 198  
 Db 320 REIRTIASDPDERFFNNV 337

RESULT 2  
 AAB64657  
 ID AAB64657 standard; protein; 1152 AA.

XX AAB64657;

XX 22-MAR-2001 (first entry)

XX Human secreted protein BLAST search protein SEQ ID NO: 167.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.

OS Homo sapiens.

XX WO2000077197-A1.

XX 21-DEC-2000.

XX 01-JUN-2000; 2000WO-US014934.

XX 11-JUN-1999; 99US-0138599P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (ROSE/) ROSEN C A.  
 XX Rosen CA, Ruben SM, Komatsoulis GA;  
 PI WPI; 2001-032312/04.

XX Isolated nucleic acid molecule encoding a human secreted protein is used  
 PT in preventing, treating or ameliorating a medical condition.

PS Disclosure; Page 543-546; 558pp; English.

XX The invention relates to the isolation of genes AAF32757-F32803 encoding  
 CC the human secreted proteins AAB64549-B64594. The sequence is a search  
 CC result from a BLASTX homology search. The genes and proteins are useful  
 CC for preventing, ameliorating or treating medical conditions, e.g. by  
 CC protein or gene therapy. The genes are isolated from a range of human  
 CC tissues disclosed in the specification. The nucleic acids, proteins,  
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment and  
 CC prevention of: (a) cancer, e.g. breast and ovarian cancer, and other  
 CC cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal  
 CC tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's  
 CC disease, allergies, autoimmune haemolytic anaemia, autoimmune  
 CC thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,  
 CC rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders  
 CC such as myocardial ischaemias; (d) wound healing; (e) neurological  
 CC diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases  
 CC such as viral, bacterial, fungal and parasitic infections

XX Sequence 1152 AA;

Query Match 100.0%; Score 1025; DB 4; Length 1152;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-108;  
 Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYPMSVEVQTLRLVGLKFLDP 60  
 Db 140 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYPMSVEVQTLRLVGLKFLDP 199  
 Qy 61 EQIQVGLVQGESPVHWSLGDFTKKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQ 120  
 Db 200 EQIQVGLVQGESPVHWSLGDFTKKEEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQ 259  
 Qy 121 SHGGRPEAARLLVVVTDGESHGDELPALAKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180  
 Db 260 SHGGRPEAARLLVVVTDGESHGDELPALAKACEAGRVTRYGIAVLGHYLRQRDPSSFL 319  
 Qy 181 REIRTIASDPDERFFNNV 198  
 Db 320 REIRTIASDPDERFFNNV 337

RESULT 3  
 AAB64658  
 ID AAB64658 standard; protein; 1152 AA.

XX AAB64658;

XX 22-MAR-2001 (first entry)

XX Human secreted protein BLAST search protein SEQ ID NO: 168.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.

OS Homo sapiens.

XX WO2000077197-A1.

XX 21-DEC-2000.

```
XX 01-JUN-2000; 2000WO-US014934.
XX
XX 11-JUN-1999; 99US-0138599P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM, Komatsoulis GA;
XX WPI; 2001-032312/04.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is used
XX in preventing, treating or ameliorating a medical condition.
XX
XX Disclosure; Page 547-551; 558pp; English.
XX
XX The invention relates to the isolation of genes AAF32757-F32803 encoding
XX the human secreted proteins AAB64549-B64594. The sequence is used as a
XX query sequence for doing BLASTX searches to identify homologous
XX sequences. The genes and proteins are useful for preventing, ameliorating
XX or treating medical conditions, e.g. by protein or gene therapy. The
XX genes are isolated from a range of human tissues disclosed in the
XX specification. The nucleic acids, proteins, antibodies and (ant)agonists
XX are useful in the diagnosis, treatment and prevention of: (a) cancer,
XX e.g. breast and ovarian cancer, and other cancers of the adrenal gland,
XX bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
XX urogenital; (b) immune disorders e.g. Addison's disease, allergies,
XX autoimmune hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
XX Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
XX wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections
XX
XX Sequence 1152 AA;
XX
XX Query Match 100.0%; Score 1025; DB 4; Length 1152;
XX Best Local Similarity 100.0%; Pred. No. 2.5e-108;
XX Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 GICARVDASFOQGS LAPTAQRCPTYMDVIVL DGSNIYPSWSEVQVFLRLVGLKLFIDP 60
XX |||||
XX 140 GICARVDASFOQGS LAPTAQRCPTYMDVIVL DGSNIYPSWSEVQVFLRLVGLKLFIDP 199
XX |||||
XX
XX 61 EQIQVGLVQGESPVHEWSLGDFTKSEVVRAAKNLSRREGRETAKAQMVACTEGFSQ 120
XX |||||
XX 200 EQIQVGLVQGESPVHEWSLGDFTKSEVVRAAKNLSRREGRETAKAQMVACTEGFSQ 259
XX |||||
XX
XX 121 SHGGRPEAARLLVVVT DGS HDGELPAAL KACEAGRVTRYGI AVLGHYLRQRDPSSFL 180
XX |||||
XX 260 SHGGRPEAARLLVVVT DGS HDGELPAAL KACEAGRVTRYGI AVLGHYLRQRDPSSFL 319
XX |||||
XX
XX 181 REIRTIASDPERFFENV 198
XX |||||
XX 320 REIRTIASDPERFFENV 337
XX |||||
XX
XX RESULT 4
XX AAY32242
XX ID AAY32242 standard; protein; 1167 AA.
XX
XX AC AAY32242;
XX
XX 15-FEB-2000 (first entry)
XX
XX Human integrin subunit alpha-10.
XX
XX Integrin alpha-10; Isal0; human; trauma; rheumatoid arthritis;
XX osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;
XX therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker.
XX
XX Homo sapiens.
XX
```

```
XX Key Location/Qualifiers
XX Peptide 1..22
XX /note= "signal peptide"
XX Protein 23..1145
XX /note= "mature protein"
XX Domain 23..1120
XX /note= "extracellular domain"
XX Modified-site 98
XX /note= "N-glycosylated"
XX Domain 162..359
XX /note= "I-domain"
XX Modified-site 336
XX /note= "N-glycosylated"
XX Modified-site 364
XX /note= "N-glycosylated"
XX Binding-site 494..502
XX /note= "cation binding site motif"
XX Binding-site 558..566
XX /note= "cation binding site motif"
XX Binding-site 620..628
XX /note= "cation binding site motif"
XX Modified-site 733
XX /note= "N-glycosylated"
XX Modified-site 839
XX /note= "N-glycosylated"
XX Modified-site 921
XX /note= "N-glycosylated"
XX Modified-site 1018
XX /note= "N-glycosylated"
XX Modified-site 1039
XX /note= "N-glycosylated"
XX Domain 1121..1145
XX /note= "transmembrane domain"
XX Domain 1122..1167
XX /note= "cytoplasmic domain, specifically claimed in Claim
XX 21"
XX
XX WO9951639-A1.
XX
XX 14-OCT-1999.
XX
XX 31-MAR-1999; 99WO-SB0000544.
XX
XX 02-APR-1998; 98SE-00001164.
XX 28-JAN-1999; 99SE-00000319.
XX
XX (ACTI-) ACTIVE BIOTECH AB.
XX
XX Lundgren-Akerlund E;
XX WPI; 2000-052639/04.
XX N-PSDB; AAZ34719.
XX
XX New isolated integrin subunit alpha-10, used as a marker or target
XX molecule for cells during development, regeneration and pathological
XX conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
XX inflammation.
XX
XX Claim 1; Fig 6; 90pp; English.
XX
XX This sequence represents novel human chondrocyte integrin subunit alpha-
XX 10 (Isal0). A splice variant is given in AAY32243. The invention relates
XX to a recombinant or isolated integrin heterodimer comprising the alpha10
XX subunit in association with subunit beta (especially beta-1). The
XX heterodimer and the subunit alpha-10 can be used as markers or targets of
XX all types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts.
XX They can also be used: for treating pathological conditions involving
XX Isal0, such as damage to cartilage, trauma, rheumatoid arthritis or
XX osteoarthritis; for detecting the formation of cartilage during embryonal
XX development, physiological or therapeutic repair of cartilage, or
XX detecting regeneration of cartilage or chondrocytes during
XX transplantation of cartilage or chondrocytes; for selection and analysis
```



CC or for sorting, isolating or purification of chondrocytes and for in  
 CC vitro studies of differentiation of chondrocytes; and as a target for  
 CC anti-adhesive drugs or molecules in tendon, ligament, skeletal muscle or  
 CC other tissues where adhesion impairs the function of the tissue (all  
 CC claimed). ISa10 binding entities can be used to determine the  
 CC differentiation-state of cells during embryonic development, angiogenesis  
 CC or development of cancer, in pathological conditions such as rheumatoid  
 CC arthritis, osteoarthritis or cancer, in tissue regeneration or in  
 CC therapeutic and physiological repair of cartilage (claimed). A  
 CC vaccine comprising the integrin heterodimer or subunit alpha-10 is also  
 CC claimed. ISa10 polynucleotides, vectors, host cells and methods of  
 CC producing recombinant ISa10 are also claimed

XX SQ Sequence 1167 AA;

Query Match 100.0%; Score 1025; DB 3; Length 1167;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-108; Mismatches 0; Indels 0; Gaps 0;  
 Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDP 60  
 |||||  
 Db 140 GICARVDASFOQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDP 199  
 |||||  
 Qy 61 EQIQVGLVQGESPVHWSLGDPTKBEVVRAAKNLSRREGRETQTAQIMVACTEGFSQ 120  
 |||||  
 Db 200 EQIQVGLVQGESPVHWSLGDPTKBEVVRAAKNLSRREGRETQTAQIMVACTEGFSQ 259  
 |||||  
 Qy 121 SHGGRPEARLLVVVTGESHGDBELPAALKACAGRVTRYGVIAVLGHYLRORDPSFL 180  
 |||||  
 Db 260 SHGGRPEARLLVVVTGESHGDBELPAALKACAGRVTRYGVIAVLGHYLRORDPSFL 319  
 |||||  
 Qy 181 REIRTIASDPDERFFNV 198  
 |||||  
 Db 320 REIRTIASDPDERFFNV 337  
 |||||

RESULT 5

AAB64584  
 ID AAB64584 standard; protein; 1167 AA.

XX AC AAB64584;

XX DT 22-MAR-2001 (first entry)

XX DE Human secreted protein #37.

XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antitumor;  
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.

XX OS Homo sapiens.

XX XX WO200077197-A1.

XX XX 21-DEC-2000.

XX PF 01-JUN-2000; 2000WO-US014934.

XX PR 11-JUN-1999; 99US-0138599P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PA (ROSE/) ROSEN C A.

XX PI Rosen CA, Ruben SM, Komatsoulis GA;

XX DR WPI; 2001-032312/04.

XX DR N-PSDB; AAF32793.

XX PT Isolated nucleic acid molecule encoding a human secreted protein is used  
 PT in preventing, treating or ameliorating a medical condition.

XX

PS Claim 11; Page 496-500; 558pp; English.

XX Sequences AAB64549-B64594 represent the amino acid sequences of 47 human  
 CC secreted proteins encoded by the genes AAF32757-P32803. The genes and  
 CC proteins are useful for preventing, ameliorating or treating medical  
 CC conditions, e.g. by protein or gene therapy. The genes are isolated from  
 CC a range of human tissues disclosed in the specification. The nucleic  
 CC acids, proteins, antibodies and (ant)agonists are useful in the  
 CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
 CC ovarian cancer, and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b)  
 CC immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
 CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
 CC infectious diseases such as viral, bacterial, fungal and parasitic  
 CC infections

XX SQ Sequence 1167 AA;

Query Match 100.0%; Score 1025; DB 4; Length 1167;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-108; Mismatches 0; Indels 0; Gaps 0;  
 Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDP 60

Db 140 GICARVDASFOQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDP 199

Qy 61 EQIQVGLVQGESPVHWSLGDPTKBEVVRAAKNLSRREGRETQTAQIMVACTEGFSQ 120

Db 200 EQIQVGLVQGESPVHWSLGDPTKBEVVRAAKNLSRREGRETQTAQIMVACTEGFSQ 259

Qy 121 SHGGRPEARLLVVVTGESHGDBELPAALKACAGRVTRYGVIAVLGHYLRORDPSFL 180

Db 260 SHGGRPEARLLVVVTGESHGDBELPAALKACAGRVTRYGVIAVLGHYLRORDPSFL 319

Qy 181 REIRTIASDPDERFFNV 198

Db 320 REIRTIASDPDERFFNV 337

RESULT 6

ABP99490  
 ID ABP99490 standard; protein; 1167 AA.

XX AC ABP99490;

XX DT 26-MAR-2003 (first entry)

XX DE Human secreted protein SEQ ID NO 434.

XX KW Human; secreted protein; nootropic; neuroprotective; cytostatic;  
 KW virucide; dermatological; immunosuppressive; antinflammatory; anti-HIV;  
 KW vulnary; antibacterial; antiparkinsonian; antitickling; antianaemic;  
 KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;  
 KW antinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;  
 KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;  
 KW cardiovascular disorder; neurological disease; nephrotropic;  
 KW gene therapy.

XX OS Homo sapiens.

XX XX WO200277186-A2.

XX PN 03-OCT-2002.

XX PD 26-MAR-2002; 2002WO-US009188.

XX PF 27-MAR-2001; 2001US-0278650P.

XX PR 12-SEP-2001; 2001US-00950082.

XX PR 12-SEP-2001; 2001US-00950083.

XX

```
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-040583/03.
DR N-PSDB; ABZ66911.
XX
XX New human secreted proteins encoded by genes contained in cDNA clones
PT (e.g. HGAC19), useful for preventing, treating or diagnosing e.g. AIDS,
PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
PT West Nile fever.
XX
XX Claim 1; Page 1395-1398; 2423pp; English.
XX
XX The invention relates to novel human genes (ABZ66891-ABZ68209) and the
CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections
XX
XX Sequence 1167 AA;
XX
Query Match 100.0%; Score 1025; DB 6; Length 1167;
Best Local Similarity 100.0%; Pred. No. 2.5e-108;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKLFIDP 60
Db 140 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKLFIDP 199
QY 61 EQIQVGLVQYGESPVHEWSLGDFTKEEVVRAAKNLSRRREGRETAKTAQAINVACTEGFSQ 120
Db 200 EQIQVGLVQYGESPVHEWSLGDFTKEEVVRAAKNLSRRREGRETAKTAQAINVACTEGFSQ 259
QY 121 SHGGRPEARLLVVVTDGESHGDELPALAKACAGRVTRYGVIAVLGHYLRQRDPSSFL 180
Db 260 SHGGRPEARLLVVVTDGESHGDELPALAKACAGRVTRYGVIAVLGHYLRQRDPSSFL 319
QY 181 REIRTIASDPDERFFNV 198
Db 320 REIRTIASDPDERFFNV 337
RESULT 7
ABR00964
ID ABR00964 standard; protein; 1167 AA.
XX
AC ABR00964;
XX
XX 12-MAY-2003 (first entry)
XX
XX Human gene 18-encoded secreted protein HA0AG15, SEQ ID NO:445.
XX
XX Human; secreted protein; cancer; tumour; hyperproliferative disorder;
KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
KW drug screening; chromosome identification; chromosome mapping;
KW cytostatic; gene therapy; antinflammatory; immunomodulator; anti-HIV;
KW antianemic; vulnery; chromosome 1q21.
XX
OS Homo sapiens.
XX
XX WO200277013-A2.
XX
```

```
XX
PD 03-OCT-2002.
XX
XX 26-MAR-2002; 2002WO-US009370.
XX
XX 27-MAR-2001; 2001US-0278650P.
PR 12-SEP-2001; 2001US-00950082.
PR 12-SEP-2001; 2001US-00950083.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2003-040578/03.
DR N-PSDB; ABZ73298.
XX
XX New human secreted proteins and nucleic acids, useful for detecting or
PT treating cancer or other hyperproliferative disorders, autoimmune
PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
XX
XX Claim 13; Page 1387-1390; 2474pp; English.
XX
XX ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
CC invention also encompasses antibodies specific for the secreted proteins,
CC the use of the secreted proteins in drug screening and recombinant
CC vectors and host cells comprising a nucleic acid of the invention. The
CC secreted proteins are thought to be involved in biological activities
CC associated with cellular signalling, cellular differentiation, cell
CC migration, prohormone activation and neurotransmitter activity. The
CC secreted proteins, nucleic acids encoding them, antibodies or antibody
CC fragments specific for the secreted proteins, and modulators of protein
CC activity are useful for diagnosing or treating cancers or other
CC hyperproliferative disorders. Additionally, the secreted proteins and
CC their nucleic acids may also be used in the treatment of autoimmune
CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
CC wound healing. Nucleic acids of the invention may be used for chromosome
CC identification, chromosome mapping, in gene therapy, for identifying
CC individuals from minute biological samples, as hybridisation probes, and
CC as molecular weight markers. The present sequence represents a human
CC secreted protein of the invention
XX
XX Sequence 1167 AA;
XX
Query Match 100.0%; Score 1025; DB 6; Length 1167;
Best Local Similarity 100.0%; Pred. No. 2.5e-108;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKLFIDP 60
Db 140 GICARVDASFPQGSGLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKLFIDP 199
QY 61 EQIQVGLVQYGESPVHEWSLGDFTKEEVVRAAKNLSRRREGRETAKTAQAINVACTEGFSQ 120
Db 200 EQIQVGLVQYGESPVHEWSLGDFTKEEVVRAAKNLSRRREGRETAKTAQAINVACTEGFSQ 259
QY 121 SHGGRPEARLLVVVTDGESHGDELPALAKACAGRVTRYGVIAVLGHYLRQRDPSSFL 180
Db 260 SHGGRPEARLLVVVTDGESHGDELPALAKACAGRVTRYGVIAVLGHYLRQRDPSSFL 319
QY 181 REIRTIASDPDERFFNV 198
Db 320 REIRTIASDPDERFFNV 337
RESULT 8
ADA44026
ID ADA44026 standard; protein; 1167 AA.
XX
XX ADA44026;
XX
```



DE Human integrin alpha subunit Alpha 10 variant A domain.

XX Human; integrin alpha subunit; A domain; Alpha 10; integrin beta subunit;

KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;

KW ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;

KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy;

KW mitein.

XX Homo sapiens.

OS Synthetic.

OS

XX Key Location/Qualifiers

FT Misc-difference 193

FT /note= "Wild-type Ile substituted by any other amino

FT acid"

XX WO200209737-A1.

XX

XX 07-FEB-2002.

XX 31-JUL-2001; 2001WO-US023957.

XX 31-JUL-2000; 2000US-0221950P.

PR 11-JAN-2001; 2001US-00758493.

PR 13-MAR-2001; 2001US-00805354.

XX (GEO ) GEN HOSPITAL CORP.

PA

XX Arnaout AM, Li R, Xiong J;

PI WPI; 2002-188687/24.

DR

XX Novel high affinity integrin polypeptide useful for treating restenosis

PT and parasitic diseases, comprises all or part of variant integrin alpha

PT subunit A domain or variant integrin beta subunit A-like domain.

XX

PS Claim 53; Page; 55pp; English.

XX

CC The invention relates to a high affinity integrin polypeptide comprising

CC all or part of a variant integrin alpha subunit A domain or a variant

CC integrin beta subunit A-like domain. The polypeptide, preferably the

CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by

CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or

CC V at residue 315 and A at residue 320 have been replaced by C, is useful

CC for determining if a test compound is a candidate compound for binding to

CC CD1b or for treating an inflammatory disorder, by contacting a test

CC compound with the polypeptide and determining if the test compound binds

CC to the polypeptide. The integrin subunits are useful for reducing

CC skeletal muscle injury, for treating disorders caused by ischaemia-

CC reperfusion injury, immune complexes, restenosis and parasitic diseases,

CC to purify variant integrin polypeptide ligands and as bait proteins in

CC two-hybrid or three-hybrid assays. This sequence represents a human

CC integrin alpha subunit Alpha 10 variant A domain. Note: This variant

CC sequence is not featured in the specification but has been derived from

CC the wild-type protein shown in AAU76853

XX

SQ Sequence 195 AA;

Query Match 89.0%; Score 912; DB 5; Length 195;

Best Local Similarity 100.0%; Pred. No. 2e-96;

Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 CPTMDVIVILDGNSIYPWSEVQTFLLRLVGKLFIDPEQIQVGLVQYGSPVHWSLGD 82

Db 1 CPTMDVIVILDGNSIYPWSEVQTFLLRLVGKLFIDPEQIQVGLVQYGSPVHWSLGD 60

Qy 83 FRTKEEVVRAAKNLSRREGRTKTAQAIMVACTGFGSQSHGGRPEARLLVVVTDGSHD 142

Db 61 FRTKEEVVRAAKNLSRREGRTKTAQAIMVACTGFGSQSHGGRPEARLLVVVTDGSHD 120

Qy 143 GEELPAALKACEAGRVTRYGVIAVLGHVLRQRDPSSFLREIRTTASPDPRFFENV 198

Db 121 GEELPAALKACEAGRVTRYGVIAVLGHVLRQRDPSSFLREIRTTASPDPRFFENV 176

RESULT 11

AAU76853

ID AAU76853 standard; protein; 195 AA.

XX

AC AAU76853;

XX

DT 21-MAY-2002 (first entry)

XX

XX Human integrin alpha subunit Alpha 10 A domain.

DE

XX Human; integrin alpha subunit; A domain; Alpha 10; integrin beta subunit;

KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;

KW ischaemia-reperfusion injury; immune complex; parasitic disease;

KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy.

XX

OS Homo sapiens.

XX

PN WO200209737-A1.

XX

PD 07-FEB-2002.

XX

XX 31-JUL-2001; 2001WO-US023957.

XX

XX 31-JUL-2000; 2000US-0221950P.

PR 11-JAN-2001; 2001US-00758493.

PR 13-MAR-2001; 2001US-00805354.

XX (GEO ) GEN HOSPITAL CORP.

PA

XX Arnaout AM, Li R, Xiong J;

PI WPI; 2002-188687/24.

DR

XX Novel high affinity integrin polypeptide useful for treating restenosis

PT and parasitic diseases, comprises all or part of variant integrin alpha

PT subunit A domain or variant integrin beta subunit A-like domain.

XX

PS Example 2; Fig 5; 55pp; English.

XX

CC The invention relates to a high affinity integrin polypeptide comprising

CC all or part of a variant integrin alpha subunit A domain or a variant

CC integrin beta subunit A-like domain. The polypeptide, preferably the

CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by

CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or

CC V at residue 315 and A at residue 320 have been replaced by C, is useful

CC for determining if a test compound is a candidate compound for binding to

CC CD1b or for treating an inflammatory disorder, by contacting a test

CC compound with the polypeptide and determining if the test compound binds

CC to the polypeptide. The integrin subunits are useful for reducing

CC skeletal muscle injury, for treating disorders caused by ischaemia-

CC reperfusion injury, immune complexes, restenosis and parasitic diseases,

CC to purify variant integrin polypeptide ligands and as bait proteins in

CC two-hybrid or three-hybrid assays. This sequence represents the human

CC integrin alpha subunit Alpha 10 A domain

XX

SQ Sequence 195 AA;

Query Match 89.0%; Score 912; DB 5; Length 195;

Best Local Similarity 100.0%; Pred. No. 2e-96;

Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 CPTMDVIVILDGNSIYPWSEVQTFLLRLVGKLFIDPEQIQVGLVQYGSPVHWSLGD 82

Db 1 CPTMDVIVILDGNSIYPWSEVQTFLLRLVGKLFIDPEQIQVGLVQYGSPVHWSLGD 60

Qy 83 FRTKEEVVRAAKNLSRREGRTKTAQAIMVACTGFGSQSHGGRPEARLLVVVTDGSHD 142

Db 61 FRTKEEVVRAAKNLSRREGRTKTAQAIMVACTGFGSQSHGGRPEARLLVVVTDGSHD 120

Qy 143 GEELPAALKACEAGRVTRYGVIAVLGHVLRQRDPSSFLREIRTTASPDPRFFENV 198

Db 121 GEELPAALKACEAGRVTRYGVIAVLGHVLRQRDPSSFLREIRTTASPDPRFFENV 176

Db 121 GEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDRFFPNV 176

RESULT 12

ADR23227

ID ADR23227 standard; protein; 198 AA.

XX

AC ADR23227;

XX

DT 04-NOV-2004 (first entry)

XX

DE Human integrin alpha 10 subunit A domain.

XX

XX Human; integrin; inflammation; antiinflammatory; vasotropic.

XX

OS Homo sapiens.

XX

XX

XX Key Location/Qualifiers

XX Domain 1..195

XX FT /label = A\_domain

XX FT Misc-difference 193

XX FT /note= "Invariant Ile residue"

XX

XX WO2004066914-A2.

XX

XX 12-AUG-2004.

XX

XX 12-MAY-2003; 2003WO-US014919.

XX

XX 10-MAY-2002; 2002US-00144259.

XX

XX (GSHO ) GEN HOSPITAL CORP.

XX

XX Arnaut MA, Li R, Xiong J;

XX

XX WPI; 2004-593980/57.

DR GENBANK; XP\_002097.

XX

XX Novel variant integrin CD11b alpha subunit polypeptide, useful for

XX determining candidate compound for binding to CD11b, and for determining

XX compound as activation-dependent ligand.

XX

XX Disclosure; SEQ ID NO 7; 128pp; English.

XX

XX The present sequence is that of the human integrin alpha 10 subunit A

XX domain. This includes an invariant Ile residue. The invention features

XX variant integrin alpha subunit polypeptides in which the invariant Ile is

XX substituted by Gly, Ala or some other amino acid (e.g. Val) or is

XX deleted. The polypeptide can include part or all of the A domain.

XX Replacing the invariant Ile creates a variant integrin polypeptide that

XX is more active (i.e. in solution has a greater proportion of ligand-

XX forming polypeptides) than the wild-type form of the subunit. Variant

XX integrin polypeptides of the invention are useful in assays for compounds

XX that bind to a variant ligand, that interfere with or enhance the binding

XX of an integrin ligand to integrin, and for identifying antibodies-

XX specific ligands. They are also useful for generating antibodies, e.g.

XX monoclonal antibodies, which bind to the high efficiency form of an

XX integrin. Some such antibodies recognise an epitope that is either not

XX present or not accessible on an integrin that is in a lower affinity

XX conformation. The invention also provides methods of administering a

XX variant integrin polypeptide, or an antibody that selectivity binds it,

XX to identify a ligand which binds to an active integrin. Such assays are

XX useful for diagnosing inflammation, e.g. occult inflammation (e.g.

XX abscess or an active arteriosclerotic lesion). Variant integrin

XX polypeptides can also be used to affect the bioavailability of a variant

XX integrin polypeptide ligand and to treat disorders associated with

XX aberrant or unwanted integrin expression or activity, such as vascular

XX injury.

XX

XX Sequence 198 AA;

XX

Query Match 89.0%; Score 912; DB 8; Length 198;

Best Local Similarity 100.0%; Pred. No. 2.1e-96;

Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 CPTYMDVIVLDGNSIYPMSSEVQTFLLRLVGLKLFIDPEIQVLQVYGSPVHWSLGD 82

Db 1 CPTYMDVIVLDGNSIYPMSSEVQTFLLRLVGLKLFIDPEIQVLQVYGSPVHWSLGD 60

QY 83 FRTKEEVVRAAKNLSRRREGRETAKTAQIMVACTEGFSQSHGSRPEAARLLVVVTDGESH 142

Db 61 FRTKEEVVRAAKNLSRRREGRETAKTAQIMVACTEGFSQSHGSRPEAARLLVVVTDGESH 120

QY 143 GEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDRFFPNV 198

Db 121 GEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDRFFPNV 176

RESULT 13

ADR41424

ID ADR41424 standard; protein; 1049 AA.

XX

AC ADR41424;

XX

DT 07-OCT-2004 (first entry)

XX

XX Human CD-like molecule HAOAD02, SEQ ID NO:223.

DE

XX

XX Human; CD-like molecule; cluster of differentiation; diagnosis;

XX prevention; immune disorder; immunodeficiency; autoimmune disorder;

XX blood-related disorder; haematological disorder; haemostatic disorder;

XX thrombotic disorder; hyperproliferative disorder; cancer; tumour;

XX apoptotic disorder; cardiovascular disorder; respiratory disorder;

XX angiogenic disorder; neovascularisation; neurological disorder;

XX endocrine disorder; reproductive system disorder; infectious disease;

XX gastrointestinal disorder; drug screening; tissue regeneration;

XX chemotaxis; gene therapy; antibody therapy; drug targeting;

XX chromosome mapping; forensic analysis; immunophenotyping; cytostatic;

XX haemostatic; tranquiliser; vulnery; antiinflammatory; nephrotropic;

XX cardiant; antiallergic; anti-HIV; antirheumatic; antiarthritic;

XX antipsoratic; immunosuppressive; vasotropic; neurotropic; neuroprotective;

XX antithyroid; thyromimetic; gynaecological; virucide; hepatotropic;

XX antibacterial; dermatological; chromosome 1q21.

XX

OS Homo sapiens.

XX

XX WO200226930-A2.

XX

XX 04-APR-2002.

XX

XX 25-SEP-2001; 2001WO-US029838.

XX

XX 26-SEP-2000; 2000US-0235484P.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

XX

XX Rosen CA, Birse CE;

PI

XX WPI; 2002-405050/43.

XX

XX N-PSDB; ADR41248.

XX

XX Novel polynucleotides and polypeptides useful for treating, preventing or

XX ameliorating cardiovascular, renal, neurovascular, and autoimmune

XX disorders.

XX

XX Claim 11; SEQ ID NO 223; 1243pp; English.

XX

XX The invention relates to 167 novel human CD (cluster of differentiation)-

XX like molecules (ADR41388-ADR41563) and to cDNAs encoding them (seqid:11)-

XX

XX Sequence 1049 AA;

XX

Query Match 82.0%; Score 840.5; DB 5; Length 1049;

Best Local Similarity 86.6%; Pred. No. 4.3e-87;

Matches 168; Conservative 5; Mismatches 14; Indels 7; Gaps 2;

QY 12 PQGSLAPTAQR-----CPTYMDV---VIVLDGNSIYPWSEVQTFRLRLVGKLFIDPQIQ 64  
Db 26 PCGCFIPASGKPGFHCPTLNIHSCCHCLGWLHSIYPWSEVQTFRLRLVGKLFIDPQIQ 85  
QY 65 VGLVOYGESPVHEWSLGDFTKKEVRAAKNLSRREGRETKTAQAIWVACTEGFSQSHGG 124  
Db 86 VGLVOYGESPVHEWSLGDFTKKEVRAAKNLSRREGRETKTAQAIWVACTEGFSQSHGG 145  
QY 125 RPEARLLVVVTGESHGDSHGLPAAKACBAGRVTRYGIAVLGHYLRQRDPSSFLREIR 184  
Db 146 RPEARLLVVVTGESHGDSHGLPAAKACBAGRVTRYGIAVLGHYLRQRDPSSFLREIR 205  
QY 185 TIASDDPDERFFNV 198  
Db 206 TIASDDPDERFFNV 219

RESULT 14

AAV32282  
ID AAY32282 standard; protein; 303 AA.

XX AC AAY32282;

XX DT 15-FEB-2000 (first entry)

XX DE Mouse integrin subunit alpha-10 partial sequence.

XX KW Integrin alpha-10; Isal0; mouse; trauma; rheumatoid arthritis;  
KW osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;  
KW therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker.  
XX OS Mus musculus.

XX FH Key Location/Qualifiers  
XX PT Misc-difference 15 /note= "encoded by TWG"

XX PN W09951639-A1.

XX PD 14-OCT-1999.

XX PF 31-MAR-1999; 99WO-SE000544.

XX PR 02-APR-1998; 98SE-00001164.

XX PR 28-JAN-1999; 99SE-00000319.

XX PA (ACT1-) ACTIVE BIOTECH AB.

XX PI Lundgren-Akerlund E;

XX WPI; 2000-052639/04.

XX N-PSDB; AAZ34835.

XX New isolated integrin subunit alpha-10, used as a marker or target molecule for cells during development, regeneration and pathological conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or inflammation.

XX Example 15; Fig 15; 90pp; English.

XX This sequence represents a polypeptide encoded by an isolated partial genomic DNA (see AAZ34835) of murine integrin subunit alpha-10 (Isal0). The mouse genomic alpha-10 sequence was used to generate a targeting vector for knockout experiments. The invention relates to a recombinant or isolated integrin heterodimer comprising the alpha10 subunit in association with subunit beta (especially beta-1). The heterodimer and the subunit alpha-10 can be used as markers or targets of all types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts. They can also be used for treating pathological conditions involving Isal0, such as: damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis; for detecting the formation of cartilage during embryonal development, physiological or therapeutic repair of cartilage, or detecting regeneration of cartilage or chondrocytes during transplantation of

CC cartilage or chondrocytes; for selection and analysis or for sorting,  
CC isolating or purification of chondrocytes and for in vitro studies of  
CC differentiation of chondrocytes; and as a target for anti-adhesive drugs  
CC or molecules in tendon, ligament, skeletal muscle or other tissues where  
CC adhesion impairs the function of the tissue (all claimed). Binding  
CC entities, polynucleotides (PNS) or oligonucleotides (ONS) can be used for  
CC e.g. determining the differentiation-state of cells during embryonic  
CC development, angiogenesis or development of cancer, in pathological  
CC conditions such as rheumatoid arthritis, osteoarthritis or cancer, in  
CC tissue regeneration or in therapeutic and physiological repair of  
CC cartilage (claimed). The PNS and ONS can also be used for detecting  
CC pathological conditions such as atherosclerosis or inflammation  
CC (claimed). A vaccine comprising the integrin heterodimer or subunit alpha  
CC -10 is also claimed. Vectors, host cells and methods of producing  
CC recombinant Isal0 are provided  
XX SQ Sequence 303 AA;

Query Match 77.8%; Score 797.5; DB 3; Length 303;

Best Local Similarity 95.1%; Pred. No. 6.5e-83;

Matches 156; Conservative 5; Mismatches 2; Indels 1; Gaps 1;

QY 1 GICARVDASFPQGS LAPTAQRCP TYMDV VIVLDGNSIYPWSEVQTFRLRLVGKLFIDP 60

Db 141 GICARVDASFPQGS LAPTAQRCP TYMDV VIVLDGNSIYPWSEVQTFRLRLVGKLFIDP 200

QY 61 EQIQVGLVQYGESPVHEWSLGDFTKKEVRAAKNLSRREGRETKTAQAIWVACTEGFSQ 120

Db 201 EQIQVGLVQYGESPVHEWSLGDFTKKEVRAAKNLSRREGRETKTAQAIWVACTEGFSQ 259

QY 121 SHGGRPEARLLVVVTGESHGDSHGLPAAKACBAGRVTRYGIA 164

Db 260 SHGGRPEARLLVVVTGESHGDSHGLPAAKACBAGRVTRYGIA 303

RESULT 15

AA50087  
ID AA50087 standard; protein; 1188 AA.

XX AC AA50087;

XX DT 19-MAR-2001 (first entry)

XX DE Murine A259.

XX KW Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;  
KW kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;  
KW rheumatoid arthritis.

XX OS Mus sp.

XX FH Key Location/Qualifiers

XX PT Domain 1..1141

XX PT /label= Extracellular\_domain

XX PT Peptide 1..22

XX PT /label= Signal\_peptide

XX PT Protein 23..1188

XX PT /label= Mature\_protein

XX PT Domain 39..74

XX PT /label= Integrin\_alphasubunit\_repeat\_domain\_#1

XX PT Domain 115..157

XX PT /label= Integrin\_alphasubunit\_repeat\_domain\_#2

XX PT Domain 164..345

XX PT /label= I\_domain

XX PT Domain 367..392

XX PT /label= Integrin\_alphasubunit\_repeat\_domain\_#3

XX PT Domain 421..455

XX PT /label= Integrin\_alphasubunit\_repeat\_domain\_#4

XX PT Domain 478..516

XX PT /label= Integrin\_alphasubunit\_repeat\_domain\_#5

XX PT Domain 540..575

XX PT /label= Integrin\_alphasubunit\_repeat\_domain\_#6

XX PT Domain 602..640





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:17:21 ; Search time 24.7306 Seconds  
(without alignments)  
2658.060 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337

Perfect score: 1025

Sequence: 1 GICARVDASFPQGS LAPTA.....FLREIRTIASDPDERFFNV 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1418010 seqs, 331997259 residues

Total number of hits satisfying chosen parameters: 1418010

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1025	100.0	1167	16	US-10-741-601-531
2	1025	100.0	1177	16	US-10-741-601-532
3	912	89.0	195	10	US-09-805-354-7
4	912	89.0	195	11	US-09-758-493-7
5	912	89.0	195	14	US-10-144-259-7
6	544	53.1	707	9	US-09-764-870-313
7	544	53.1	707	14	US-10-125-540-313
8	544	53.1	1034	10	US-09-984-130-43
9	544	53.1	1034	10	US-09-836-353A-43
10	544	53.1	1188	15	US-10-291-265-338
11	544	53.1	1188	15	US-10-291-265-810
12	544	53.1	1189	10	US-09-984-130-35
13	544	53.1	1189	10	US-09-836-353A-35

14	544	53.1	1189	15	US-10-262-839-4	Sequence 4, Appli
15	541	52.8	437	15	US-10-108-260A-3386	Sequence 3386, Ap
16	492	48.0	193	10	US-09-805-354-8	Sequence 8, Appli
17	492	48.0	193	11	US-09-758-493-8	Sequence 8, Appli
18	492	48.0	193	14	US-10-144-259-8	Sequence 8, Appli
19	492	48.0	1120	15	US-10-262-839-6	Sequence 6, Appli
20	472.5	46.1	1177	17	US-10-741-600-1161	Sequence 1161, Ap
21	467.5	45.6	1151	10	US-09-984-130-103	Sequence 103, App
22	467.5	45.6	1151	10	US-09-836-353A-103	Sequence 103, App
23	467.5	45.6	1179	10	US-09-918-715-250	Sequence 250, App
24	465.5	45.4	1180	10	US-09-918-715-307	Sequence 307, App
25	452.5	44.1	214	13	US-10-061-658-5	Sequence 5, Appli
26	452.5	44.1	214	15	US-10-625-260-5	Sequence 5, Appli
27	452.5	44.1	214	15	US-10-474-832-63	Sequence 63, Appli
28	449.5	43.9	214	13	US-10-061-658-6	Sequence 6, Appli
29	449.5	43.9	214	13	US-10-061-658-9	Sequence 9, Appli
30	449.5	43.9	214	15	US-10-625-260-6	Sequence 6, Appli
31	449.5	43.9	214	15	US-10-625-260-9	Sequence 9, Appli
32	449.5	43.9	214	15	US-10-474-832-64	Sequence 64, Appli
33	441	43.0	1147	15	US-10-336-603A-42	Sequence 42, Appli
34	441	43.0	1148	16	US-10-872-198-147	Sequence 147, App
35	441	43.0	1181	14	US-10-160-354-2	Sequence 2, Appli
36	441	43.0	1181	15	US-10-295-027-1286	Sequence 1286, Ap
37	441	43.0	1181	15	US-10-211-462-187	Sequence 187, App
38	433.5	42.3	212	9	US-09-996-738-5	Sequence 5, Appli
39	430.5	42.0	212	9	US-09-996-738-6	Sequence 6, Appli
40	427	41.7	195	10	US-09-805-354-5	Sequence 5, Appli
41	427	41.7	195	11	US-09-758-493-5	Sequence 5, Appli
42	427	41.7	195	14	US-10-144-259-5	Sequence 5, Appli
43	424	41.4	192	15	US-10-474-832-59	Sequence 59, Appli
44	424	41.4	192	15	US-10-474-832-61	Sequence 61, Appli
45	419	40.9	192	15	US-10-346-863-57	Sequence 57, Appli

#### ALIGNMENTS

RESULT 1  
US-10-741-601-531  
; Sequence 531, Application US/10741601  
; Publication No. US20040166519A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; FILE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01500  
; CURRENT APPLICATION NUMBER: US/10/741,601  
; CURRENT FILING DATE: 2003-12-22  
; NUMBER OF SEQ ID NOS: 26415  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 531  
; LENGTH: 1167  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-741-601-531

Query Match 100.0%; Score 1025; DB 16; Length 1167;  
Best Local Similarity 100.0%; Pred. No. 1.8e-100;  
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GICARVDASFPQGS LAPTAQRCPTVMDVIVLGDNSIYPWSEVOTFLRLVGLFIDP	60
Db	140	GICARVDASFPQGS LAPTAQRCPTVMDVIVLGDNSIYPWSEVOTFLRLVGLFIDP	199
Qy	61	EQIQVGLVQGESVPVHWSLGDFTKEVRAAKNLSRREGRETAKTAQAINVACTEGFSQ	120
Db	200	EQIQVGLVQGESVPVHWSLGDFTKEVRAAKNLSRREGRETAKTAQAINVACTEGFSQ	259
Qy	121	SHGGRPEARLLVVVTDGESHGDELPAALKACAGRVTRYGVIAVLHYLRQRDPSSFL	180
Db	260	SHGGRPEARLLVVVTDGESHGDELPAALKACAGRVTRYGVIAVLHYLRQRDPSSFL	319
Qy	181	REIRTIASDPDERFFNV	198

```
Db 320 REIRTIASDPDERFFNV 337
|||||
RESULT 2
US-10-741-601-532
; Sequence 532, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C0001500
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 532
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-741-601-532

Query Match 100.0%; Score 1025; DB 16; Length 1177;
Best Local Similarity 100.0%; Pred. No. 1.9e-100;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GICARVDASFOQSLAPTAQRCPTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDP 60
|||||
Db 140 GICARVDASFOQSLAPTAQRCPTYMDVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDP 199
|||||
Qy 61 EQIQVGLVQYGESPVHWSLGDFTKBEVVRAAKNLSRREGRETAKQAIMVACTEGFSQ 120
|||||
Db 200 EQIQVGLVQYGESPVHWSLGDFTKBEVVRAAKNLSRREGRETAKQAIMVACTEGFSQ 259
|||||
Qy 121 SHGGRPEARLLVVVTGESHGDBELPAALKACAGRVTRYGIAVLGHLRQRDPSSFL 180
|||||
Db 260 SHGGRPEARLLVVVTGESHGDBELPAALKACAGRVTRYGIAVLGHLRQRDPSSFL 319
|||||
Qy 181 REIRTIASDPDERFFNV 198
|||||
Db 320 REIRTIASDPDERFFNV 337
|||||

RESULT 3
US-09-805-354-7
; Sequence 7, Application US/09805354
; Publication No. US20030078375A1
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amin
; APPLICANT: Li, Rui
; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-536001
; CURRENT APPLICATION NUMBER: US/09/805,354
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US 09/758,493
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-805-354-7

Query Match 89.0%; Score 912; DB 10; Length 195;
Best Local Similarity 100.0%; Pred. No. 2.1e-89;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 CPTYMDVVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDPEQIQVGLVQYGESPVHWSLGD 82
|||||
Db 1 CPTYMDVVIVLDGNSIYPWSEVQTFLLRLVGLKFLIDPEQIQVGLVQYGESPVHWSLGD 60
|||||
Qy 83 FRTKEEVVRAAKNLSRREGRETAKQAIMVACTEGFSQSHGGRPEARLLVVVTGESH 142
|||||
Db 61 FRTKEEVVRAAKNLSRREGRETAKQAIMVACTEGFSQSHGGRPEARLLVVVTGESH 120
|||||
Qy 143 GEELPAALKACEAGRVTRYGIAVLGHLRQRDPSSFLREIRTIASDPDERFFNV 198
|||||
Db 121 GEELPAALKACEAGRVTRYGIAVLGHLRQRDPSSFLREIRTIASDPDERFFNV 176
|||||

US-09-647-544-2_copy_140_337.rapb
```

```
; ORGANISM: Homo sapiens
US-10-144-259-7

Query Match      89.0%; Score 912; DB 14; Length 195;
Best Local Similarity 100.0%; Pred. No. 2.1e-89;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 CPTMDVIVLDGNSIYPSEVOTFLRLVGLKFLIDPEQIOVLGVQYGESPVHWSLGD 82
Db 1 CPTMDVIVLDGNSIYPSEVOTFLRLVGLKFLIDPEQIOVLGVQYGESPVHWSLGD 60

Qy 83 FRTKEEVRAAKNLSRREGRTKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTDGESH 142
Db 61 FRTKEEVRAAKNLSRREGRTKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTDGESH 120

Qy 143 GEELPAALKACEAGRTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDERFFNV 198
Db 121 GEELPAALKACEAGRTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDERFFNV 176

RESULT 6
US-09-764-870-313
; Sequence 313, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ14
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 313
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-870-313

Query Match      53.1%; Score 544; DB 9; Length 707;
Best Local Similarity 53.0%; Pred. No. 4e-49;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

Qy 1 GICARVDASFQPGSLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFRLRLVGLKFLIDP 60
Db 153 GMC SRVNSNFRFSKTVAPALQRCOTYMDIVIVLDGNSIYPWVEVQHFLINILKKFYIGP 212

Qy 61 EQIQVGLVQYGESPVHWSLGDERTKEEVRAAKNLSRREGRTKTAQAIMVACTEGFSQ 120
Db 213 GQIQVGVVQYGEDVHFEHLNDYRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQK 272

Qy 121 SHGGRPEAARLLVVVTDGESHGDELPAAKACACAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 273 --GGRKGAKKVMIVITDGHSDSPDLKVIQQSERDNTVRYAVAVLGYNRRGINPETFL 330

; REIRTIASDPDERFFNV 198
; NEIKYIASDPDDKHFFNV 348

RESULT 7
US-10-125-540-313
; Sequence 313, Application US/10125540
; Publication No. US20030059875A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ14C1
; CURRENT APPLICATION NUMBER: US/10/125,540
; CURRENT FILING DATE: 2002-04-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 313
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-540-313

Query Match      53.1%; Score 544; DB 14; Length 707;
Best Local Similarity 53.0%; Pred. No. 4e-49;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

Qy 1 GICARVDASFQPGSLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFRLRLVGLKFLIDP 60
Db 153 GMC SRVNSNFRFSKTVAPALQRCOTYMDIVIVLDGNSIYPWVEVQHFLINILKKFYIGP 212

Qy 61 EQIQVGLVQYGESPVHWSLGDERTKEEVRAAKNLSRREGRTKTAQAIMVACTEGFSQ 120
Db 213 GQIQVGVVQYGEDVHFEHLNDYRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQK 272

Qy 121 SHGGRPEAARLLVVVTDGESHGDELPAAKACACAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 273 --GGRKGAKKVMIVITDGHSDSPDLKVIQQSERDNTVRYAVAVLGYNRRGINPETFL 330

; REIRTIASDPDERFFNV 198
; NEIKYIASDPDDKHFFNV 348

RESULT 8
US-09-984-130-43
; Sequence 43, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-43

Query Match      53.1%; Score 544; DB 10; Length 1034;
Best Local Similarity 53.0%; Pred. No. 6.8e-49;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

Qy 1 GICARVDASFQPGSLAPTAQRCTYMDVIVLDGNSIYPWSEVQTFRLRLVGLKFLIDP 60
Db 137 GMC SRVNSNFRFSKTVAPALQRCOTYMDIVIVLDGNSIYPWVEVQHFLINILKKFYIGP 196

Qy 61 EQIQVGLVQYGESPVHWSLGDERTKEEVRAAKNLSRREGRTKTAQAIMVACTEGFSQ 120
Db 197 GQIQVGVVQYGEDVHFEHLNDYRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQK 256

Qy 121 SHGGRPEAARLLVVVTDGESHGDELPAAKACACAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 257 --GGRKGAKKVMIVITDGHSDSPDLKVIQQSERDNTVRYAVAVLGYNRRGINPETFL 314

; REIRTIASDPDERFFNV 198
; NEIKYIASDPDDKHFFNV 348
```

```
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 9
; Sequence 43, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-353A-43

Query Match 53.1%; Score 544; DB 10; Length 1034;
Best Local Similarity 53.0%; Pred. No. 6.8e-49;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVVIVLDGNSIYPWSEVQTFLLRLVGLKLFIDP 60
Db 137 GMC SRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYPWSEVQHFLLNLIKFFYIGP 196

Qy 61 EQIQVGLVQYGESPVHWSLGDFTKKEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQ 120
Db 197 GOIQGVVQYGEDVVFHFLNDYRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQK 256

Qy 121 SHGGRPEAARLLVVVTDGSHDGEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 257 --GGRGAKKVMIVITDGHSDSPDLEKVIQOQSRDNTVRYAVAVLGYNNRRGINPETFL 314

Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 11
US-10-291-265-810
; Sequence 810, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 810
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-810

Query Match 53.1%; Score 544; DB 15; Length 1188;
Best Local Similarity 53.0%; Pred. No. 8.2e-49;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVVIVLDGNSIYPWSEVQTFLLRLVGLKLFIDP 60
Db 137 GMC SRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYPWSEVQHFLLNLIKFFYIGP 196

Qy 61 EQIQVGLVQYGESPVHWSLGDFTKKEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQ 120
Db 197 GOIQGVVQYGEDVVFHFLNDYRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQK 256

Qy 121 SHGGRPEAARLLVVVTDGSHDGEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 257 --GGRGAKKVMIVITDGHSDSPDLEKVIQOQSRDNTVRYAVAVLGYNNRRGINPETFL 314

Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 12
US-09-984-130-35

Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 10
US-10-291-265-338
; Sequence 338, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 338
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-338

Query Match 53.1%; Score 544; DB 10; Length 1034;
Best Local Similarity 53.0%; Pred. No. 6.8e-49;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVVIVLDGNSIYPWSEVQTFLLRLVGLKLFIDP 60
Db 137 GMC SRVNSNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYPWSEVQHFLLNLIKFFYIGP 196

Qy 61 EQIQVGLVQYGESPVHWSLGDFTKKEVVRAAKNLSRRREGRTKTAQAIMVACTEGFSQ 120
Db 197 GOIQGVVQYGEDVVFHFLNDYRSVKDVVEAASHIEQRGGTETRTAFGIEFARSEAFQK 256

Qy 121 SHGGRPEAARLLVVVTDGSHDGEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 257 --GGRGAKKVMIVITDGHSDSPDLEKVIQOQSRDNTVRYAVAVLGYNNRRGINPETFL 314

Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332
```

Matches	105;	Conservative	39;	Mismatches	52;	Indels	2;	Gaps	1;
Qy	1	GICARVDASFPQGS	LAPTACRCPTYMDVVIVLDGNSIYPWSEVQTFLRLRVGKLFIDP	60					
Dd	137	GMCSRNVNFRFSKTAPALQR	COTYMDIVIVLDGNSIYPWVEVOHFLINILKFFVIGP	196					
Qy	61	EOTQVLGVOCSPVHEWS	IGDTRTKEEVRAAKNLRRREGRTKTAAAIMVACTEGFSQ	120					
Dd	197	GOIQGVGVQGEDVHFH	LNDYRSVKDVVEAASHIEQRGGTETRTAFGEFARSEAFQK	256					
Qy	121	SHGGRPEAAELLVVVT	DGESHDCBELPAALKAACEAGRTVRYGIYAVLCHYLRRORDPSSFL	180					
Dd	257	--GGRCAKKVMIVTD	GSHDSPDLKVKIQSERDNVTRYAYAVLVGYNNRGINPETFL	314					
Qy	181	REINTIASDPDRFFFN	V 198						
Dd	315	NEIKYIASDDDKHFN	V 332						
RESULT 14									
US-10-262-839--4									
; Sequence 4, Application US/10262839									
; Publication No. US20040038877A1									
; GENERAL INFORMATION:									
; APPLICANT: Alsbrook, John,									
; APPLICANT: Anderson, David W.,									
; APPLICANT: Boldog, Ferenc,									
; APPLICANT: Burgess, Catherine,									
; APPLICANT: Catterton, Elina,									
; APPLICANT: Edinger, Shomit,									
; APPLICANT: Ellerman, Karen,									
; APPLICANT: Gerlach, Valerie,									
; APPLICANT: Gorman, Linda,									
; APPLICANT: Guo, Xiaojia,									
; APPLICANT: Ji, Weizhen,									
; APPLICANT: Kekuda, Ramesh,									
; APPLICANT: Leach, Martin,									
; APPLICANT: Li, Li,									
; APPLICANT: Miller, Charles,									
; APPLICANT: Patturajan, Meera,									
; APPLICANT: Reiger, Daniel,									
; APPLICANT: Rothenberg, Mark,									
; APPLICANT: Shimkets, Richard,									
; APPLICANT: Smithson, Glennda,									
; APPLICANT: Spytek, Kimberly,									
; APPLICANT: Taupier, Raymond, jr.,									
; APPLICANT: Vernet, Corine,									
; APPLICANT: Voss, Edward,									
; APPLICANT: Zerhusen, Brian,									
; APPLICANT: Zhong, Mei									
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AL									
; FILE REFERENCE: 21402-462A									
; CURRENT APPLICATION NUMBER: US/10/262,839									
; PRIOR FILING DATE: 2002-10-01									
; PRIOR APPLICATION NUMBER: 60/326,483									
; PRIOR FILING DATE: 2001-10-02									
; PRIOR APPLICATION NUMBER: 60/327,917									
; PRIOR FILING DATE: 2001-10-09									
; PRIOR APPLICATION NUMBER: 60/328,029									
; PRIOR FILING DATE: 2001-10-09									
; PRIOR APPLICATION NUMBER: 60/328,056									
; PRIOR FILING DATE: 2001-10-09									
; PRIOR APPLICATION NUMBER: 60/381,101									
; PRIOR FILING DATE: 2002-05-16									
; PRIOR APPLICATION NUMBER: 60/371,972									
; PRIOR FILING DATE: 2002-04-12									
; PRIOR APPLICATION NUMBER: 60/327,342									
; PRIOR FILING DATE: 2001-10-05									
; PRIOR APPLICATION NUMBER: 60/328,044									
; PRIOR FILING DATE: 2001-10-09									
; PRIOR APPLICATION NUMBER: 60/328,849									
; PRIOR FILING DATE: 2001-10-12									
; PRIOR APPLICATION NUMBER: 60/374,738									

```
; PRIOR FILING DATE: 2002-04-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 4
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-262-839-4

Query Match          53.1%; Score 544; DB 15; Length 1189;
Best Local Similarity 53.0%; Pred. No. 8.2e-45;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

Qy 1 GICARVDASFPQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDP 60
Db 137 GMCSRVSNSFRFSKTVAPALQRCQTYMDIVIVLDGNSIYPWVEVQHFLINILKFFYIGP 196

Qy 61 EQIQVGLVQYGESPVHWSLGDPRTKKEEVVRAAKNLSRREGRETCTAQAIMVACTEGFSQ 120
Db 197 GQIQGVVQYGEDVVFHFLNDYRSVKDVVEAASHIEQRGTETRTAFGIEFARSEAFQK 256

Qy 121 SHGGRPEAARLLVVVTGESHGDSGELPAALKACAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 257 --GGRKGAKKVMIVITDGHSDSPDLEKVIQQSERDNTTRYAVAVLGYNNRGINPETFL 314

Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 15
US-10-108-260A-3386
; Sequence 3386, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3386
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3386

Query Match          52.8%; Score 541; DB 15; Length 437;
Best Local Similarity 52.5%; Pred. No. 4.3e-49;
Matches 104; Conservative 39; Mismatches 53; Indels 2; Gaps 1;

Qy 1 GICARVDASFPQGS LAPTAQRCTYMDVIVLDGNSIYPWSEVQTFLLRLVGKLFIDP 60
Db 137 GMCSRVSNSFRFSKTVAPALQRCQTYMDIVIVLDGNSIYPWVEVQHFLINILKFFYIGP 196

Qy 61 EQIQVGLVQYGESPVHWSLGDPRTKKEEVVRAAKNLSRREGRETCTAQAIMVACTEGFSQ 120
Db 197 GQIQGVVQYGEDVVFHFLNDYRSVKDVVEAASHIEQRGTETRTAFGIEFARSEAFQK 256

Qy 121 SHGGRPEAARLLVVVTGESHGDSGELPAALKACAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 257 --GGRKGAKKVMIVITDGHSDSPDLEKVIQQSERDNTTRYAVAVLGYNNRGINPETFL 314

Qy 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332
```

Search completed: April 6, 2005, 12:53:02  
Job time : 25.7306 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:52 ; Search time 8.45027 Seconds  
(without alignments)  
1749.117 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_140\_337  
Perfect score: 1025  
Sequence: 1 GICARVDASFPQGS LAPTA.....FLREIRTIASDPDRFFNV 198

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pap.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pap.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pap.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pap.\*
- 5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pap.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	544	53.1	1217	4	US-09-949-016-7892
2	465.5	45.4	1180	4	US-09-000-004A-2
3	452.5	44.1	214	4	US-10-061-658-5
4	449.5	43.9	214	4	US-10-061-658-6
5	449.5	43.9	214	4	US-10-061-658-9
6	441	43.0	1181	4	US-09-000-004A-4
7	441	43.0	1181	4	US-09-949-016-6189
8	441	43.0	1195	4	US-09-949-016-10747
9	420.5	41.0	1183	4	US-09-532-310B-6
10	396	38.6	1183	4	US-09-532-310B-5
11	205	20.0	496	1	US-08-462-128-37
12	205	20.0	496	1	US-08-463-180-37
13	205	20.0	496	2	US-08-001-078A-1
14	205	20.0	496	2	US-08-897-443-4
15	205	20.0	496	2	US-08-463-218-1
16	205	20.0	496	4	US-09-949-016-11306
17	205	20.0	496	5	PCT-US94-00253-1
18	198	19.3	413	1	US-08-485-618-101
19	198	19.3	413	2	US-08-605-672-101
20	198	19.3	413	2	US-08-482-293A-101
21	198	19.3	413	2	US-08-943-363-101
22	198	19.3	413	3	US-09-193-043-101
23	198	19.3	413	4	US-09-688-307A-101
24	198	19.3	413	4	US-09-350-259-101
25	196	19.1	956	2	US-08-897-443-3
26	191	18.6	638	2	US-08-897-443-1
27	191	18.6	915	4	US-09-907-794A-34

28	191	18.6	915	4	US-09-905-125A-34	Sequence 34, Appl
29	191	18.6	915	4	US-09-902-775A-34	Sequence 34, Appl
30	191	18.6	915	4	US-09-906-700-34	Sequence 34, Appl
31	191	18.6	915	4	US-09-903-603A-34	Sequence 34, Appl
32	191	18.6	915	4	US-09-904-920A-34	Sequence 34, Appl
33	191	18.6	915	4	US-09-909-064A-34	Sequence 34, Appl
34	191	18.6	915	4	US-09-905-381A-34	Sequence 34, Appl
35	191	18.6	915	4	US-09-906-618-34	Sequence 34, Appl
36	191	18.6	956	4	US-09-949-016-6215	Sequence 6215, Ap
37	191	18.6	963	4	US-09-949-016-11519	Sequence 11519, A
38	191	18.6	963	4	US-09-949-016-11520	Sequence 11520, A
39	185.5	18.1	584	4	US-09-949-016-10340	Sequence 10340, A
40	185.5	18.1	584	4	US-09-949-016-10341	Sequence 10341, A
41	177.5	17.3	1178	1	US-08-199-776-2	Sequence 2, Appli
42	177.5	17.3	1178	3	US-08-663-731-2	Sequence 2, Appli
43	177.5	17.3	1178	3	US-08-879-338-2	Sequence 2, Appli
44	177.5	17.3	1178	5	PCT-US95-02044-2	Sequence 2, Appli
45	177.5	17.3	1179	4	US-09-293-238B-2	Sequence 2, Appli

## ALIGNMENTS

## RESULT 1

US-09-949-016-7892  
; Sequence 7892, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7892  
; LENGTH: 1217  
; TYPE: PRT  
; ORGANISM: Human  
; US-09-949-016-7892

Query Match	53.1%	Score 544;	DB 4;	Length 1217;
Best Local Similarity	53.0%	Pred. No. 3.9e-59;		
Matches 105;	Conservative 39;	Mismatches 52;	Indels 2;	Gaps 1;
Qy	1	GICARVDASFOQGS LAPTAQRCPTVMDVIVLDGNSIYPWSEVQTFELRLVLKFLIDP	60	
Db	166	GMCSRVSNNFRFSKTVAPALQRCQTDIVIVLDGNSIYPWVEVQHFLINILKFFYGP	225	
Qy	61	EQIQVGLVQYGESPVHWSGLDFTKSEVRAAKNLSRREGTKTAQAINVACTEGFSQ	120	
Db	226	GOIQGVVQYGEDVVHFLNDYRSVKDVEAASHIEQRCGGTETRTAFGIEFARSEAFQK	285	
Qy	121	SHGRPEARALLVVVTDGSHDGEPLPAALKACACAGRTVRYGIVLGHVLRQDRPSSFL	180	
Db	286	--GGKGGAKKVMIVITDGSHSDPLEKVIQOOSRDNVTRYAVAVLGYNNRRGINPETFL	343	
Qy	181	REIRTIASDPDRFFNV	198	
Db	344	NEIKIASDPDKHFFNV	361	

## RESULT 2

US-09-000-004A-2  
; Sequence 2, Application US/09000004A  
; Patent No. 6780603





; CURRENT APPLICATION NUMBER: US/10/061,658  
; CURRENT FILING DATE: 2002-02-01  
; PRIOR APPLICATION NUMBER: 60/137,214  
; PRIOR FILING DATE: 1999-06-01  
; PRIOR APPLICATION NUMBER: 60/130,847  
; PRIOR FILING DATE: 1999-04-22  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 214  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-10-061-658-9

Query Match 43.9%; Score 449.5; DB 4; Length 214;  
Best Local Similarity 46.1%; Pred. No. 2.4e-48;  
Matches 89; Conservative 35; Mismatches 68; Indels 1; Gaps 1;  
Qy 6 VDASFQPGSLAPTAQRCPYMDVIVLDGNSIYPWSEVQTFRLRLVGLKLFIDPEQIQV 65  
Db 1 VSPFTQVNSIAP-VQSCSTQDIDIVLDGNSIYPWSDVTAFLNDLLKRMIDIGPKQTQV 59  
Qy 66 GLVGYSPVHESLIGPRTKEEVVRAAKNLSRREGRETTKTAQAIMVACTEGFSQSHGGR 125  
Db 60 GIVGYGENVTHEFLNLYKSYSTEELVAAKTIQVGRGRTWTALGTDTRKEAFTEARGAR 119  
Qy 126 PEARLLVVVTDGSHGDEELPAALKAACEAGRVTRYGIAVLGHYLRQRDPSSFLREIRT 185  
Db 120 RGVKVMVIVTDGSHDNHRLKKVIQDCDENIQRFSAIILGNSYRGNLSKEVEBIKS 179  
Qy 186 IASDPDERFFNNV 198  
Db 180 IASEPTEKHFFNV 192

RESULT 6  
US-09-000-004A-4  
; Sequence 4, Application US/090000004A  
; Patent No. 6780603  
; GENERAL INFORMATION:  
; APPLICANT: Tslibary, Photini-Effie  
; APPLICANT: Charonia, Aristidis S.  
; APPLICANT: Setty, Suman  
; APPLICANT: Mauer, Michael  
; TITLE OF INVENTION: ANALYSIS OF ALPHA INTEGRINS FOR THE DIAGNOSIS OF DIABETIC NEPHRO  
; FILE REFERENCE: 600.314USWO  
; CURRENT APPLICATION NUMBER: US/09/000,004A  
; CURRENT FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: US 60/001,387  
; PRIOR FILING DATE: 1995-07-21  
; PRIOR APPLICATION NUMBER: US 60/001,861  
; PRIOR FILING DATE: 1995-08-03  
; PRIOR APPLICATION NUMBER: US 60/016,700  
; PRIOR FILING DATE: 1996-05-02  
; PRIOR APPLICATION NUMBER: PCT/US96/12067  
; PRIOR FILING DATE: 1996-07-19  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 1181  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-000-004A-4

Query Match 43.0%; Score 441; DB 4; Length 1181;  
Best Local Similarity 44.4%; Pred. No. 4.2e-46;  
Matches 88; Conservative 33; Mismatches 77; Indels 0; Gaps 0;  
Qy 1 GICARVDASFQPGSLAPTAQRCPYMDVIVLDGNSIYPWSEVQTFRLRLVGLKLFIDP 60  
Db 147 GVCSDISPDFQLSASFSPATQPCSLIDVVVDCDSNIYPWDVAKNLFKVFQGLDIGP 206  
Qy 61 EQIQVGLVQGESPVHESLIGDFTKEEVVRAAKNLSRREGRETTKTAQAIMVACTEGFSQ 120

Db 207 TKTQVGLIQYANNPRVFNLTNTYKTEEMIVATSTQSYGGDLTNTGAIQYARKYASA 266  
Qy 121 SHGGRPEARLLVVVTDGSHGDEELPAALKAACEAGRVTRYGIAVLGHYLRQRDPSSFL 180  
Db 267 ASGGRRSATKVMVIVTDGSHDGMKKAVIDQCNDHNLRFAGIAGVLGYNLRNALDITKNLI 326  
Qy 181 REIRTIASDPDERFFNNV 198  
Db 327 KEIKAIASIPTRYFFNV 344

RESULT 7  
US-09-949-016-6189  
; Sequence 6189, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6189  
; LENGTH: 1181  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-6189

Query Match 43.0%; Score 441; DB 4; Length 1181;  
Best Local Similarity 44.4%; Pred. No. 4.2e-46;  
Matches 88; Conservative 33; Mismatches 77; Indels 0; Gaps 0;  
Qy 1 GICARVDASFQPGSLAPTAQRCPYMDVIVLDGNSIYPWSEVQTFRLRLVGLKLFIDP 60  
Db 147 GVCSDISPDFQLSASFSPATQPCSLIDVVVDCDSNIYPWDVAKNLFKVFQGLDIGP 206  
Qy 61 EQIQVGLVQGESPVHESLIGDFTKEEVVRAAKNLSRREGRETTKTAQAIMVACTEGFSQ 120  
Db 207 TKTQVGLIQYANNPRVFNLTNTYKTEEMIVATSTQSYGGDLTNTGAIQYARKYASA 266  
Qy 121 SHGGRPEARLLVVVTDGSHGDEELPAALKAACEAGRVTRYGIAVLGHYLRQRDPSSFL 180  
Db 267 ASGGRRSATKVMVIVTDGSHDGMKKAVIDQCNDHNLRFAGIAGVLGYNLRNALDITKNLI 326  
Qy 181 REIRTIASDPDERFFNNV 198  
Db 327 KEIKAIASIPTRYFFNV 344

RESULT 8  
US-09-949-016-10747  
; Sequence 10747, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03

```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10747
; LENGTH: 1195
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10747

Query Match      43.0%; Score 441; DB 4; Length 1195;
Best Local Similarity 44.4%; Pred. No. 4.3e-46;
Matches 88; Conservative 33; Mismatches 77; Indels 0; Gaps 0;

Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVVLVDGNSIYPWSEVQTFLLRLVGLKLFIDP 60
Db 161 GVCSDISPDQLSASPSAPATPCPSLIDVVVVDGNSIYPWDVAKNFKLVQGLDGP 220

Qy 61 EQIQVGLVQGESPVHWSLGDPRTKKEEVVRAAKNLRRREGRETKTAAQAIMVACTEGFSQ 120
Db 221 TKTQVGLIQYANPRVFNLTYYKKEIMIVATQTSQYGGDLTNTFGAIVGLYLRNLDTKNLI 340

Qy 121 SHGCRPEAARLLVVVTDGSHDGEELPAALKAACEAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 281 ASGRRSATKVMVVVTDGSHDGMKXAVIDQCNDHNLIRPGIAVLGLYLRNLDTKNLI 340

Qy 181 REIRTIASDPDERFFNV 198
Db 341 KEIKAIASIPTERYFNV 358

RESULT 9
US-09-532-310B-6
; Sequence 6, Application US/09532310B
; Patent No. 6596276
; GENERAL INFORMATION:
; APPLICANT: Senger, Donald R
; Detmar, Michael
; Claffey, Kevin P
; TITLE OF INVENTION: Method for inhibiting tumor
; angiogenesis in a living subject
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David Prashker, Esq.
; STREET: P.O. Box 5387
; CITY: Magnolia
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 01930
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage
; COMPUTER: Dell PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Microsoft Word version 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/532,310B
; FILING DATE: 22-Mar-2000
; CLASSIFICATION: Unknown
; ATTORNEY/AGENT INFORMATION:
; NAME: David Prashker, Esq.
; REGISTRATION NUMBER: 29,693
; REFERENCE/DOCKET NUMBER: BIS-036
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (978) 525-3794
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1183 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-532-310B-6

Query Match      38.6%; Score 396; DB 4; Length 1183;
Best Local Similarity 42.9%; Pred. No. 2.1e-40;
Matches 85; Conservative 33; Mismatches 78; Indels 2; Gaps 2;

Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVVLVDGNSIYPWSEVQTFLLRLVGLKLFIDP 60
Db 148 GICSDVSPTFQVNVNAP-VQECSTXLDIVLVDGNSIYPMDSV-TALNDLLKRMIDGP 205
```

```
Query Match      41.0%; Score 420.5; DB 4; Length 1183;
Best Local Similarity 43.2%; Pred. No. 1.7e-43;
Matches 86; Conservative 33; Mismatches 79; Indels 1; Gaps 1;

Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVVLVDGNSIYPWSEVQTFLLRLVGLKLFIDP 60
Db 147 GVCSDISPDQLSASPSAPATPCPSLIDVVVVDGNSIYPWDVAKNFKLVQGLDGP 206

Qy 61 EQIQVGLVQGESPVHWSLGDPRTKKEEVVRAAKNLRRREGRETKTAAQAIMVACTEGFSQ 120
Db 207 TKTQVGLIQYANPRVFNLTYYKKEIMIVATQTSQYGGDLTNTFGAIVGLYLRNLDTKNLI 326

Qy 121 SHGCRPEAARLLVVVTDGSHDGEELPAALKAACEAGRVTRYGIAVLGHYLRQRDPSS-SF 179
Db 267 ASGRRSATKVMVVVTDGSHDGMKXAVIDQCNDHNLIRPGIAVLGLYLRNLDTKNLI 326

Qy 180 REIRTIASDPDERFFNV 198
Db 327 KEIKAIASIPTERYFNV 345

RESULT 10
US-09-532-310B-5
; Sequence 5, Application US/09532310B
; Patent No. 6596276
; GENERAL INFORMATION:
; APPLICANT: Senger, Donald R
; Detmar, Michael
; Claffey, Kevin P
; TITLE OF INVENTION: Method for inhibiting tumor
; angiogenesis in a living subject
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David Prashker, Esq.
; STREET: P.O. Box 5387
; CITY: Magnolia
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 01930
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage
; COMPUTER: Dell PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Microsoft Word version 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/532,310B
; FILING DATE: 22-Mar-2000
; CLASSIFICATION: Unknown
; ATTORNEY/AGENT INFORMATION:
; NAME: David Prashker, Esq.
; REGISTRATION NUMBER: 29,693
; REFERENCE/DOCKET NUMBER: BIS-036
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (978) 525-3794
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1183 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-532-310B-5

Query Match      38.6%; Score 396; DB 4; Length 1183;
Best Local Similarity 42.9%; Pred. No. 2.1e-40;
Matches 85; Conservative 33; Mismatches 78; Indels 2; Gaps 2;

Qy 1 GICARVDASFOQGS LAPTAQRCTYMDVVLVDGNSIYPWSEVQTFLLRLVGLKLFIDP 60
Db 148 GICSDVSPTFQVNVNAP-VQECSTXLDIVLVDGNSIYPMDSV-TALNDLLKRMIDGP 205
```

```

Qy 61 EQIOVGLVOYGESPVHWSLGDFTKEEVVRAAKNLSRRREGRETAKTAQAIMVACTEGFSQ 120
Db 206 KXTXGVIVXYGENTHFNLNKYSSTSEVLVAAKKIVXRGGRXTWTALGTDTARKEAFTE 265
Qy 121 SHGRPEAARLLVVVTGESHGDSGELPAALKACAGRVTRYGIAVLGHYLRQRDPSSFL 180
Db 266 ARGARGVKVMVITVDEGHDHNRLLKKVIGDCEDENIXRFSIALLOSYNRGNLSTEFV 325
Qy 181 REIRTIASDDPDERFFNV 198
Db 326 EEKSIASEPTEKFFNV 343

RESULT 11
US-08-462-128-37
; Sequence 37, Application US/08462128
; Patent No. 5686059
; GENERAL INFORMATION:
; APPLICANT: Goetinck, Paul F.
; APPLICANT: Tondravi, Mehrdad
; TITLE OF INVENTION: CARTILAGE MATRIX PROTEIN AND METHODS FOR
; TITLE OF INVENTION: USE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,128
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/006,096
; FILING DATE: 15-JAN-1993
; APPLICATION NUMBER: US 07/866,403
; FILING DATE: 10-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Paul L.
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-462-128-37

Query Match 20.0%; Score 205; DB 1; Length 496;
Best Local Similarity 33.5%; Pred. No. 8.8e-17;
Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;

Qy 28 DWIVLDGNSIYP--WSEVQTFLLRLVGLKLFIDPEQIQVGLVOYGESPVHWSLGDFT 85
Db 275 DLVFLIDGSKSVRPENFELVKKFSIQIVDTLDVSDKLAQVGLVQYSSSVRQEPFLGRPT 334
Qy 86 KEEVVRAAKNLSRRREGRETAKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTGESHGDEE 145
Db 335 KKDIIKAAVRNMSYME-KGTWMTGAALKYLIDNSFTVSSGARPGAKQKVGIVFTDGRSQDYIN 393
Qy 146 LPAALKACEAG-RVTRYGIAVLGHYLRQRDPSSFLREIRTIASDP-DERFFF 196
Db 394 -DAAKKAADLGFKMFVGVG-----NAVEDELREIASEPVAEHYFY 433

RESULT 12
US-08-463-180-37
; Sequence 37, Application US/08463180
; Patent No. 5741670
; GENERAL INFORMATION:
; APPLICANT: Goetinck, Paul F.
; APPLICANT: Tondravi, Mehrdad
; TITLE OF INVENTION: CARTILAGE MATRIX PROTEIN AND METHODS FOR
; TITLE OF INVENTION: USE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,180
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/006,096
; FILING DATE: 15-JAN-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/866,403
; FILING DATE: 10-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Paul L.
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-005DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-463-180-37

Query Match 20.0%; Score 205; DB 1; Length 496;
Best Local Similarity 33.5%; Pred. No. 8.8e-17;
Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;

Qy 28 DWIVLDGNSIYP--WSEVQTFLLRLVGLKLFIDPEQIQVGLVOYGESPVHWSLGDFT 85
Db 275 DLVFLIDGSKSVRPENFELVKKFSIQIVDTLDVSDKLAQVGLVQYSSSVRQEPFLGRPT 334
Qy 86 KEEVVRAAKNLSRRREGRETAKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTGESHGDEE 145
Db 335 KKDIIKAAVRNMSYME-KGTWMTGAALKYLIDNSFTVSSGARPGAKQKVGIVFTDGRSQDYIN 393
Qy 146 LPAALKACEAG-RVTRYGIAVLGHYLRQRDPSSFLREIRTIASDP-DERFFF 196
Db 394 -DAAKKAADLGFKMFVGVG-----NAVEDELREIASEPVAEHYFY 433

RESULT 13
US-08-001-078A-1
; Sequence 1, Application US/08001078A
; Patent No. 5872094
; GENERAL INFORMATION:
; APPLICANT: Goetinck, Paul F.
; APPLICANT: Tondravi, M., Mehrdad

```

```
; APPLICANT: Binette, Francois
; TITLE OF INVENTION: METHODS FOR PROMOTING CARTILAGE MATRIX
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, Suite 510
; CITY: BOSTON
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/001,078A
; FILING DATE: 06-JAN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Paul L.
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MGP-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-001-078A-1

Query Match 20.0%; Score 205; DB 2; Length 496;
Best Local Similarity 33.5%; Pred. No. 8.8e-17;
Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;

Qy 28 DVVLVDGNSIYP--WSEVQTFRLRLVGLKFLIDPEIQVGLVQYGSPPVHWSLGDPRFT 85
Db 275 DLVFLDGSKSVRPENFELVKKFIQIVDTLDVSDKLAQVGLVQYSSVRQEPFLGRFHT 334

Qy 86 KEEVVRAAKNLSRRREGTKTAQAIMVACTEGFSQSHGGRPEAAARLLVVVTDGESHGDEE 145
Db 335 KKDIKAAVRNMSYNE-KGTMTGAALKYILDNSFTVSSGARPGAKVGIVFTDGRSQDYIN 393

Qy 146 LPALKACEAG-RVTRYGIAVLGHYLRQRDPSSFLREIRTIASDP-DERFFF 196
Db 394 -DAAKAKDLGFKMFAVGVG-----NAVEDELREIASEPVAEHYF 433

RESULT 14
US-08-897-443-4
; Sequence 4, Application US/08897443
; Patent No. 5981263
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Kaser, Mathew
; TITLE OF INVENTION: HUMAN MATRILIN-3
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

```
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,443
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0348 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1732121
; US-08-897-443-4

Query Match 20.0%; Score 205; DB 2; Length 496;
Best Local Similarity 33.5%; Pred. No. 8.8e-17;
Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;

Qy 28 DVVLVDGNSIYP--WSEVQTFRLRLVGLKFLIDPEIQVGLVQYGSPPVHWSLGDPRFT 85
Db 275 DLVFLDGSKSVRPENFELVKKFIQIVDTLDVSDKLAQVGLVQYSSVRQEPFLGRFHT 334

Qy 86 KEEVVRAAKNLSRRREGTKTAQAIMVACTEGFSQSHGGRPEAAARLLVVVTDGESHGDEE 145
Db 335 KKDIKAAVRNMSYNE-KGTMTGAALKYILDNSFTVSSGARPGAKVGIVFTDGRSQDYIN 393

Qy 146 LPALKACEAG-RVTRYGIAVLGHYLRQRDPSSFLREIRTIASDP-DERFFF 196
Db 394 -DAAKAKDLGFKMFAVGVG-----NAVEDELREIASEPVAEHYF 433

RESULT 15
US-08-463-218-1
; Sequence 1, Application US/08463218
; Patent No. 5986052
; GENERAL INFORMATION:
; APPLICANT: Goetinck, Paul F.
; APPLICANT: Tondravi, Mehrdad
; APPLICANT: Binette, Francois
; TITLE OF INVENTION: METHODS FOR PROMOTING CARTILAGE MATRIX
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, Suite 510
; CITY: BOSTON
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,218
; FILING DATE: 05-JUNE-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/001,078
; FILING DATE: 06-JAN-1993
```

ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Paul L.  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: MGP-008DV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 496 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-463-218-1

Query Match 20.0%; Score 205; DB 2; Length 496;  
Best Local Similarity 33.5%; Pred. No. 8.8e-17;  
Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;  
Qy 28 DWIVLDGNSIYP--WSEVQTELRLVGLKLFIDPEIQVGLVQYGESPVHEWSLGDPT 85  
Db 275 DLVFLIDGSKSVRPENFELVKKFSIQIVDTLDVSDKLAQVGLVQYSSVRQEFPLGRFHT 334  
Qy 86 KEEVVRAAKNLRSREGRETAKTQAIMVACTEGFSQSHGGRPEAAARLLVVVTDGESHGEE 145  
Db 335 KKDIAKAVRNNSYNE-KCTMTGAALKYLLIDNSFTVSSGARGPAOKVGIVFTDGRSQDYIN 393  
Qy 146 LPAALKACEAG-RVTRYGIAGVGLHYLRQRDPSPFLREIRTIASDP-DERFFF 196  
Db 394 -DAAKAKDLGFKMFVGVG-----NAVEDELREIASEPVAEHYFY 433

Search completed: April 6, 2005, 12:24:22  
Job time : 9.45027 secs

***This Page Blank (uspto)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 5.18011 Seconds  
(without alignment)  
3459.921 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_952\_986

Perfect score: 192

Sequence: 1 YEVHPYCTLPVGPCEPKFTLRVQNLGCVVSGLI 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trenbl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	100.0	1167	1 ITAG HUMAN	O75578 homo sapien
2	175	91.1	288	2 QBEM12	QBEM12 mus musculu
3	65.5	34.1	301	1 NODP_AZOB	P28603 azospirillu
4	59	30.7	823	2 Q8WY18	Q8WY18 homo sapien
5	59	30.7	1189	1 ITAG HUMAN	Q8WY18 homo sapien
6	58.5	30.5	2275	2 Q8QS41	Q8QS41 pongine her
7	58	30.2	444	2 Q8NZW4	Q8NZW4 brachydanio
8	58	30.2	496	2 Q7PTM2	Q7PTM2 anopheles g
9	58	30.2	497	2 Q7PTJ9	Q7PTJ9 anopheles g
10	58	30.2	695	2 Q9YI02	Q9YI02 brachydanio
11	57.5	29.9	3338	2 Q82PH6	Q82PH6 streptomyce
12	56	29.2	94	2 Q8YT51	Q8YT51 human immun
13	56	29.2	283	2 Q842M3	Q842M3 rhodococcus
14	56	29.2	404	2 Q6FSB8	Q6FSB8 candida gla
15	56	29.2	600	1 ACEK_BORPA	Q7W2P4 bordetella
16	56	29.2	823	2 Q8CEB4	Q8CEB4 mus musculu
17	56	29.2	1188	1 ITAG MOUSE	P61622 mus musculu
18	56	29.2	1188	2 Q7TQC3	Q7TQC3 mus musculu
19	55	28.6	321	2 Q6C493	Q6C493 yarrowia li
20	55	28.6	892	2 Q7WYF8	Q7WYF8 alcaligenes
21	55	28.6	1054	2 Q9SNY7	Q9SNY7 nicotiana t
22	55	28.6	1940	2 Q6GPD0	Q6GPD0 xenopus lae
23	54.5	28.4	299	1 NODP_RHISB	O07308 rhizobium s
24	54.5	28.4	302	2 Q8L3E5	Q8L3E5 xanthomonas
25	54.5	28.4	302	2 Q8P610	Q8P610 xanthomonas
26	54.5	28.4	302	2 Q8PHD0	Q8PHD0 xanthomonas
27	54.5	28.4	303	2 Q7UW11	Q7UW11 rhodospirill
28	54.5	28.4	445	2 Q7XVK9	Q7XVK9 oryza sativ
29	54.5	28.4	1173	2 Q8ZX19	Q8ZX19 pyrobaculum
30	54	28.1	215	2 Q8AGB0	Q8AGB0 human immun
31	54	28.1	217	2 Q8AGB1	Q8AGB1 human immun

#### RESULT 1

ITAG\_HUMAN STANDARD; PRT; 1167 AA.

AC O75578; Q9UHZ8;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 05-JUN-2004 (Rel. 44, Last annotation update)

DE Integrin alpha-10 precursor.

GN Name=ITGA10;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RT TISSUE=Articular chondrocytes;

RC MEDLINE=98532078; PubMed=9685391; DOI=10.1074/jbc.273.32.20383;

RA Camper L., Hellman U., Lundgren-Akerlund E.;

RT "Isolation, cloning, and sequence analysis of the integrin subunit

RT alpha10, a beta1-associated collagen binding integrin expressed on

RT chondrocytes.";

RL J. Biol. Chem. 273:20383-20389(1998).

RN [2]

RP SEQUENCE FROM N.A.

RT TISSUE=Endothelial cells, and Heart;

RC MEDLINE=20169197; PubMed=10702680;

RA Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,

RA Wang S.-X., Langley R., Krissansen G.W.;

RT "The integrin alpha10 subunit: expression pattern, partial gene

RT structure, and chromosomal localization.";

RL Cytogenet. Cell Genet. 87:238-244(1999).

CC -!- FUNCTION: Integrin alpha-10/beta-1 is a receptor for collagen.

CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-10

CC associates with beta-1.

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -!- TISSUE SPECIFICITY: Widely expressed with highest expression in

CC muscle and heart. Found in articular cartilage.

CC -!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins

CC with I-domains do not undergo protease cleavage.

CC -!- SIMILARITY: Belongs to the integrin alpha chain family.

CC -!- SIMILARITY: Contains 7 FG-GAP repeats.

CC -!- SIMILARITY: Contains 1 VWFA domain.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

EMBL; AF074015; AAC31952.1; -

EMBL; AF112345; AAF21944.1; -

EMBL; AF172723; AAF61638.1; -

DR



```

Query Match          34.1%; Score 65.5; DB 1; Length 301;
Best Local Similarity 48.4%; Pred. No. 1.2;
Matches 15; Conservative 5; Mismatches 8; Indels 3; Gaps 2;

QY      7 GTLPVGP--PFFKTLRVQNLCGYVVSGLI 35
         ||||| : : : : : : : : : : : :
DB     228 GRPLNPGTPEMR-RVRFTLGCYPLSGAI 257

RESULT 4
Q8WYI8 PRELIMINARY; PRT; 823 AA.
ID Q8WYI8 AC Q8WYI8
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
MT MSTP018.
DE GN Name=MST018;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Aorta;
RA Xu Y.Y., Sun L.Z., Wu Q.Y., Liu Y.Q., Liu B., Zhao B., Wang X.Y.,
RA Song L., Ye J.J., Sheng H., Gao Y., Zhang C.L., Zhang J., Wei Y.J.,
RA Sun R.H., Jiang Y.X., Zhao X.W., Liu S., Liu L.S., Ding J.F.,
RA Gao R.L., Qiang B.Q., Yuan J.G., Liew C.C., Zhao M.S., Hui R.T.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC CC -1 SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1 SIMILARITY: Belongs to the integrin alpha chain family.
DR EMBL; AF11799; AAL39001.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:000515; F:protein binding; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.
DR Pfam; PF01839; FG-GAP; 1.
DR PRINTS; PR01185; INTEGRINA.
DR SMART; SM00191; Int_alpha.
KW Cell adhesion; Integrin; Transmembrane.
SQ SEQUENCE 823 AA; 92672 MW; DE4E78079DCD4925 CRC64;

Query Match          30.7%; Score 59; DB 2; Length 823;
Best Local Similarity 41.7%; Pred. No. 28;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY      12 GPGEPEKTLRVQNLCGYVVSGLI 35
         ||||| : : : : : : : : : : : :
DB     616 GIGPFSCIFRIQLGLFPIHGMW 639

RESULT 5
ITAH HUMAN STANDARD; PRT; 1189 AA.
ID ITAH HUMAN AC Q9UKX5; Q9UKQ1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-11 precursor.
GN Name=ITGALL;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal heart, and Osteoblast;
RX MEDLINE=599417678; PubMed=10486209; DOI=10.1006/geno.1999.5909;
RA Lehnert K., Ni J.J., Leung E., Gough S.M., Weaver A., Yao W.P., Liu D.,
RA Wang S.-X., Morris C.N., Krissansen G.W.;
```



RT "The human cytomegalovirus genome revisited: comparison with the  
 RL chimpanzee cytomegalovirus genome."  
 J. Gen. Virol. 84:17-28(2003).

DR EMBL; AF480884; AAM00697.1; -

DR InterPro; IPR006928; Herpes teg.N.

DR Pfam; PF04843; Herpes teg.N.1.

SQ SEQUENCE 2275 AA; 255930 MW; 45BBA419CA576BCD CRC64;

Query Match 30.5%; Score 58.5; DB 2; Length 2275;

Best Local Similarity 34.0%; Pred. No. 96;

Matches 16; Conservative 7; Mismatches 7; Indels 17; Gaps 3;

Qy 1 YEHPY-----GTLVPVPG--PEFKTLRVQNLGCVV 31

Db 1680 YDLKPYFSQQQQQRAAQDGSVPLGPAQPPPEANTLLNK-LFCYV 1725

RESULT 7

ID Q6NZW4 PRELIMINARY; PRT; 444 AA.

AC Q6NZW4;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)

DE Skib protein (Fragment).

GN Name=skib;

OS Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OC NCBI\_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.W., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,

RA Jones S.J., Marra M.A.

RT "Generation and initial analysis of more than 15,000 full-length human

and mouse cDNA sequences."

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RA Strausberg R.

RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC065946; AAH65946.1; -

DR ZFIN; ZDB-GENE-990715-10; skib.

DR GO; GO:0005634; C:nucleus; IEA.

DR GO; GO:0008151; P:cell growth and/or maintenance; IEA.

DR InterPro; IPR009061; Putativ DNA\_bind.

DR InterPro; IPR010919; SAND-like.

DR InterPro; IPR003380; TransForm\_Ski.

DR Pfam; PF02437; Ski\_Sno; 1.

FT NON TER 444

SQ SEQUENCE 444 AA; 50214 MW; 31F75A5AD7C9FC66 CRC64;

Query Match 30.2%; Score 58; DB 2; Length 444;

Best Local Similarity 35.9%; Pred. No. 20;

Matches 14; Conservative 4; Mismatches 11; Indels 10; Gaps 1;

Qy 5 PYGTLPVPGPEF-----KTLRVQNLGCVVVG 33

Db 58 PAQTPVMPGFLFIPSDRSTERCETVLERITSCFVVG 96

RESULT 8

ID Q7PTM2

AC Q7PTM2 PRELIMINARY; PRT; 496 AA.

DT 01-MAR-2004 (TrEMBLrel. 26, Created)

DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE ENSANGP0000013480 (Fragment).

GN Name=ENSANGG0000010991;

OS Anopheles gambiae str. PEST.

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.

OC NCBI\_TaxID=180454;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PEST;

RA Anopheles Genome Sequencing Consortium;

RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: Belongs to the cytochrome P450 family.

CC -!- CAUTION: The sequence shown here is derived from an

EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is

preliminary data.

DR EMBL; AAB01008799; EAA03690.2; -

DR HSP; P14779; IUPZ.

DR GO; GO:0004497; F:monooxygenase activity; IEA.

DR GO; GO:0006118; P:electron transport; IEA.

DR InterPro; IPR001128; Cytochrome\_P450.

DR InterPro; IPR002401; EP4501.

DR Pfam; PF00067; P450\_1.

DR PRINTS; PR00463; EP4501.

DR PRINTS; PR00385; P450.

DR PROSITE; PS00086; CYTOCHROME\_P450; UNKNOWN\_1.

DR Heme; Monooxygenase; Oxidoreductase.

FT NON TER 1

FT NON TER 496

FT NON TER 496

SQ SEQUENCE 496 AA; 56959 MW; 7766741C5F4AED38 CRC64;

Query Match

Best Local Similarity

30.2%; Score 58; DB 2; Length 496;

26.2%; Pred. No. 22;

Matches 17; Conservative 4; Mismatches 8; Indels 36; Gaps 2;

Qy 2 EVHPYGLPVGPGP-----EPKTLRVQNL----- 26

Db 427 ERHPYAYLPFGSGPRNCIGIRYAWLSMKILIAHLVNRYPKTKLMEDLEIKFAILLRIM 486

Qy 27 -GCVV 30

Db 487 NGCLV 491

RESULT 9

ID Q7PTJ9

AC Q7PTJ9 PRELIMINARY; PRT; 497 AA.

DT 01-MAR-2004 (TrEMBLrel. 26, Created)

DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE ENSANGP0000006300 (Fragment).

GN Name=ENSANGG0000010991;

OS Anopheles gambiae str. PEST.

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.

OC NCBI\_TaxID=180454;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PEST;

```

RA Anopheles Genome Sequencing Consortium;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.
CC -I- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008799; EAA03815.2; -.
DR HSSP; P14779; 1JPZ.
DR GO; GO:0004497; P:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR002401; EP4501.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00463; EP4501.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
FT NON_TER 497 497
FT NON_TER 1
SQ SEQUENCE 497 AA; 57002 MW; CD558BFD2C9C701C CRC64;

Query Match 30.2%; Score 58; DB 2; Length 497;
Best Local Similarity 26.2%; Pred. No. 22;
Matches 17; Conservative 4; Mismatches 8; Indels 36; Gaps 2;

Qy 2 EVHPYGLPVGPGP-----EPKTLRVQNL----- 26
Db 427 ERHPYALPFGGPRNCIGIRYAWLSMKILLIAHLVRYRKYTLKMEDLEIKFAILLRIM 486

Qy 27 -GCYV 30
Db 487 NGCLV 491

RESULT 10
QyY102 PRELIMINARY; PRT; 695 AA.
AC QYV102;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Nuclear oncoprotein.
GN Name=skib; Synonyms=skib; (Danio rerio).
OS Brachydanio rerio (Zebrafish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20368164; PubMed=10906458; DOI=10.1016/S0925-4773(00)00351-8;
RA Kaufman C.D., Martinez-Rodriguez G., Hackett P.B. Jr.;
RT "Ectopic expression of c-ski disrupts gastrulation and neural
RL patterning in zebrafish.";
RL Mech. Dev. 95:147-162(2000).
DR EMBL; AF060118; AAC64707.1; -.
DR HSSP; P12755; 1MR1.
DR ZFIN; ZDB-GENE-990715-10; skib.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0008151; P:cell growth and/or maintenance; IEA.
DR InterPro; IPR009061; Putativ_DNA_bind.
DR InterPro; IPR010919; SAND like.
DR InterPro; IPR003380; Transform_Ski.
DR Pfam; PF02437; Ski_Sno1.1.
SQ SEQUENCE 695 AA; 79135 MW; 9BD424C32F61C4ED CRC64;

Query Match 30.2%; Score 58; DB 2; Length 695;
Best Local Similarity 35.9%; Pred. No. 32;
Matches 14; Conservative 4; Mismatches 11; Indels 10; Gaps 1;

Qy 5 PYGTLPGVGPGEF-----KTLRVQNLGCYVSG 33
Db 58 PAQTFFVMPGFLFIPDSRSTRCRTVLERTISCFVWG 96

```

## RESULT 11

```

Q82PH6 PRELIMINARY; PRT; 3338 AA.
ID Q82PH6
AC Q82PH6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Putative transcriptional activator SRCAP homolog.
DE OrderedLocNames=SAV926;
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680;
RX MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;
RA Omura S., Ikeda H., Iehikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis; deducing the ability of producing secondary
RT metabolites.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Iehikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005024; BAC68636.1; -.
DR InterPro; IPR000504; RNA_rec_mot.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 3338 AA; 345044 MW; 5254F9C38439054B CRC64;

Query Match 29.9%; Score 57.5; DB 2; Length 3338;
Best Local Similarity 56.5%; Pred. No. 2e+02;
Matches 13; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

Qy 8 TLPVGQPEPKTTLRVQNLGCYV 30
Db 2091 TLP-GPDPEFSPALRLRLGLQV 2112

RESULT 12
Q9Y751 PRELIMINARY; PRT; 94 AA.
ID Q9Y751
AC Q9Y751;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN Name=env;
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98445411; PubMed=9770526; DOI=10.1073/pnas.95.21.12568;
RA Markham R.B., Wang W.C., Weisstein A.E., Wang Z., Munoz A.,
RA Templeton A., Margolick J., Vlahov D., Quinn T., Farzadegan H.,
RA Yu X.P.;
RT "Patterns of HIV-1 evolution in individuals with differing rates of
RT CD4 T cell decline.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:12568-12573(1998).
DR EMBL; AF089156; AAC78866.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.

```

Genolevures:

Dujon B., Sherman D., Fischner G., Durrens P., Casaregola S., Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E., Goffard N., Frangeul L., Aigle M., Anichouard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C., Boismame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A., Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A., Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R., Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H., Nicaud J.M., Nikolski M., Ostas S., Ozier-Kalogeropoulos O., Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A., Swennene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B., Zentoul-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A., Bouchier C., Coudron B., ScarPELLI C., Gaillardin C., Weissenbach J., Wincker P., Souciet J.L.; "Genome evolution in yeasts."; Nature 430:35-44(2004).

EMBL; CR380954; CAG59609.1; GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . ; IEA. GO; GO:0006807; P:nitrogen metabolism; IEA. InterPro; IPR003010; NtIsE/CNhydts.

DR Pfam; PF00795; CN hydrolase; 1.  
DR PROSITE; PS0263; CN HYDROLASE; 1.  
SQ SEQUENCE 404 AA; 45985 MW; 2C448299FB6130D6 CRC64;

Query Match 29.2%; Score 56; DB 2; Length 404;  
Best Local Similarity 46.7%; Pred.No. 34;  
Matches 14; Conservative 1; Mismatches 13; Indels 2; Gaps 2;

QY 5 PYGTLPGVGPBKFTLRVQNLL-GCVVSG 33  
|| || | || | || | || |  
DB 68 PY-VCKAGEGPSPELAKRVSKLFGCYTVIG 96

RESULT 15  
ACEK\_BORPA STANDARD; PRF; 600 AA.  
ID Q7W2P4;  
AC 29-MAR-2004 (Rel. 43, Created)  
DT 29-MAR-2004 (Rel. 43, Last sequence update)  
DT 25-OCT-2004 (Rel. 45, Last annotation update)  
DE Isocitrate dehydrogenase kinase/phosphatase (EC 2.7.1.116)  
DE (EC 3.1.3.-) (IDH kinase/phosphatase) (IDHK/P).  
GN Names-acek; OrderedLocNames=BPP4360;  
OS Bordetella parapertussis.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Alcaligenaceae; Bordetella.  
OX NCBI\_TaxID=519;  
RX [1]  
SEQUENCE FROM N.A.  
RC STRAIN=12822 / ATCC BAA-587;  
RP MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngl227;  
RX Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R., Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L., Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A., Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I., Chillingworth T., Collins M., Cronin A., Davis P., Doggett J., Felwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K., Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C., Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K., Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K., Unwin L., Whitehead S., Barrall B.G., Maskell D.J.  
RT "Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica." Nat. Genet. 35:32-40(2003).  
RL CC -!- FUNCTION: Bifunctional enzyme which can phosphorylate or dephosphorylate isocitrate dehydrogenase (IDH) on a specific serine residue. This is a regulatory mechanism which enables bacteria to bypass the krebs cycle via the glyoxylate shunt in response to the source of carbon. When bacteria are grown on glucose, IDH is fully active and unphosphorylated, but when grown on acetate or ethanol, the activity of IDH declines drastically concomitant with its phosphorylation (by similarity!).



```

CC -|- CATALYTIC ACTIVITY: ATP + [isocitrate dehydrogenase (NADP+)] = ADP
CC + [isocitrate dehydrogenase (NADP+)] phosphate.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -|- SIMILARITY: Belongs to the aceK family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; BX640436; CAB39639.1; -.
CC HAMAP; MF_00747; -; 1.
CC InterPro; IPR010452; AceK.
CC Pfam; PF06315; AceK; 1.
CC ATP-binding; Complete proteome; Glyoxylate bypass; Hydrolase; Kinase;
CC Multifunctional enzyme; Protein phosphatase; Transferase;
CC Tricarboxylic acid cycle.
CC NP_BIND 335 341 ATP (By similarity).
CC BINDING 356 356 ATP (By similarity).
CC ACT_SITE 390 390 By similarity.
CC SEQUENCE 600 AA; 68516 MW; A4F10F6987971226 CRC64;
CC -----
Query Match 29.2%; Score 56; DB 1; Length 600;
Best Local Similarity 33.3%; Pred. No. 52;
Matches 13; Conservative 4; Mismatches 14; Indels 8; Gaps 1;

Qy 5 PYGTLVPVGPGEFKTTLRVQ-----NLGCYVVSGLI 35
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 196 PYADLPDAGPRLASDCQIQVLGSLFFRNTGAYIVGRLI 234

```

Search completed: April 6, 2005, 12:15:25  
Job time : 7.18011 secs

Result No.	Score	Query			DB	ID	Description
		Match	Length	#			
1	65.5	34.1	301	2	I39754	nodP protein - Azotobacter vinelandii	
2	54	28.1	2139	2	S46404	vitellogenin - vitellogenin - vitellogenin - vitellogenin	
3	52	27.1	101	2	S61172	hypothetical protein - hypothetical protein - hypothetical protein - hypothetical protein	
4	52	27.1	437	1	TVFVSK	transforming protein - transforming protein - transforming protein - transforming protein	
5	52	27.1	564	2	I53106	gene g11 protein - gene g11 protein - gene g11 protein - gene g11 protein	
6	52	27.1	649	2	C32575	C-ski protein FB27 - C-ski protein FB27 - C-ski protein FB27 - C-ski protein FB27	
7	52	27.1	713	2	B32575	C-ski protein FB28 - C-ski protein FB28 - C-ski protein FB28 - C-ski protein FB28	
8	52	27.1	728	1	TVHUSK	transforming protein - transforming protein - transforming protein - transforming protein	
9	52	27.1	750	2	A32575	C-ski protein FB29 - C-ski protein FB29 - C-ski protein FB29 - C-ski protein FB29	
10	51.5	26.8	390	2	T25966	hypothetical protein - hypothetical protein - hypothetical protein - hypothetical protein	
11	51.5	26.8	507	1	S22973	cytochrome P450 52 - cytochrome P450 52 - cytochrome P450 52 - cytochrome P450 52	
12	51	26.6	402	2	T37694	hypothetical protein - hypothetical protein - hypothetical protein - hypothetical protein	
13	51	26.6	890	2	A48753	NFAT transcription factor - NFAT transcription factor - NFAT transcription factor - NFAT transcription factor	
14	51	26.6	1053	2	S34172	sucrose-phosphate 6-epimerase - sucrose-phosphate 6-epimerase - sucrose-phosphate 6-epimerase - sucrose-phosphate 6-epimerase	
15	50.5	26.3	300	2	T03464	probable methylene tetrahydropteroyl transferase - probable methylene tetrahydropteroyl transferase - probable methylene tetrahydropteroyl transferase - probable methylene tetrahydropteroyl transferase	
16	50.5	26.3	317	2	F82672	ATP sulfurylase, alpha - ATP sulfurylase, alpha - ATP sulfurylase, alpha - ATP sulfurylase, alpha	
17	50	26.0	428	2	T27763	hypothetical protein - hypothetical protein - hypothetical protein - hypothetical protein	
18	49.5	25.8	348	2	D90422	alcohol dehydrogenase - alcohol dehydrogenase - alcohol dehydrogenase - alcohol dehydrogenase	
19	49.5	25.8	666	2	G81717	glycosyl hydrolase - glycosyl hydrolase - glycosyl hydrolase - glycosyl hydrolase	
20	49	25.5	136	2	C90254	hypothetical protein - hypothetical protein - hypothetical protein - hypothetical protein	
21	49	25.5	286	2	G72625	probable ribosomal protein - probable ribosomal protein - probable ribosomal protein - probable ribosomal protein	
22	49	25.5	446	2	B90463	proline/betaine transporter - proline/betaine transporter - proline/betaine transporter - proline/betaine transporter	
23	49	25.5	753	2	S58331	dolichyl-phosphate synthase - dolichyl-phosphate synthase - dolichyl-phosphate synthase - dolichyl-phosphate synthase	
24	49	25.5	771	2	T43612	transposase - transposase - transposase - transposase	
25	48.5	25.3	317	2	A12676	sulfate adenylate transferase - sulfate adenylate transferase - sulfate adenylate transferase - sulfate adenylate transferase	
26	48.5	25.3	341	2	G97458	hypothetical protein - hypothetical protein - hypothetical protein - hypothetical protein	
27	48.5	25.3	352	2	G70620	probable argC protein - probable argC protein - probable argC protein - probable argC protein	
28	48.5	25.3	505	2	F71266	probable tpr protein - probable tpr protein - probable tpr protein - probable tpr protein	
29	48.5	25.3	653	2	C70742	probable rsbU protein - probable rsbU protein - probable rsbU protein - probable rsbU protein	

RESULT 3  
S61172  
hypothetical protein YDR377w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein D9481.21  
C:Species: Saccharomyces cerevisiae  
C>Date: 23-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 09-Jul-2004  
C:Accession: S61172  
R/Ding, H.  
submitted to the EMBL Data Library, June 1995  
A:Description: The sequence of S. cerevisiae cosmid 9481.  
A:Reference number: S61172  
A:Accession: S61172  
A:Molecule type: DNA  
A:Residues: 1-101 <DIN>  
A:Cross-references: UNIPROT:Q06405; EMBL:U28373; NID:G849184; PIDN:AAB64813.1; PID:G849184  
A:Experimental source: strain S288C (AB972)  
C:Genetics:  
A:Gene: SGD:ATP17  
A:Cross-references: SGD:S0002785; MIPS:YDR377w  
A:Map position: 4R  
C:Superfamily: Saccharomyces hypothetical protein YDR377w

Query Match 27.1%; Score 52; DB 2; Length 101;  
Best Local Similarity 47.6%; Pred. No. 4.1;  
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 3 VHPYGTLPVGPPEFKTTLRV 23  
|||:|||||:  
Db 34 VHFYKSLPQGPAPAKANTRL 54  
|||:|||||:

RESULT 4  
TVPVSK  
transforming protein ski - avian erythroblastosis virus (strain Sloan-Kettering)  
C:Species: avian erythroblastosis virus  
C>Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 09-Jul-2004  
C:Accession: A32574  
R:Scavnezer, E.; Brodeur, D.; Brennan, L.A.  
Mol. Cell. Biol. 9, 4038-4045, 1989  
A:Title: The v-ski oncogene encodes a truncated set of c-ski coding exons with limited  
A:Reference number: A32574; MUID:89384634; PMID:2674685  
A:Accession: A32574  
A:Molecule type: DNA  
A:Residues: 1-437 <STA>  
A:Cross-references: UNIPROT:P17863  
C:Genetics:  
A:Gene: v-ski  
C:Superfamily: ski transforming protein  
C:Keywords: transforming protein

Query Match 27.1%; Score 52; DB 1; Length 437;  
Best Local Similarity 35.3%; Pred. No. 20;  
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

QY 10 PVGPGPEF-----KTLRVQNLGCVVWSG 33  
|||:|||||:  
Db 46 PVMPGPFMPDSRSTERCETILGETISCFVVG 79  
|||:|||||:

RESULT 5  
I53106  
gene gli protein - mouse (fragment)  
C:Species: Mus sp. (mouse)  
C>Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 23-Feb-1997  
C:Accession: I53106  
R:Walterhouse, D.; Ahmed, M.; Slusarski, D.; Kalamaras, J.; Boucher, D.; Holmgren, R.; I  
Dev. Dyn. 196, 91-102, 1993  
A:Title: gli, a zinc finger transcription factor and oncogene, is expressed during norma  
A:Reference number: I53106; MUID:93372381; PMID:8364225  
A:Accession: I53106  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA

A:Residues: 1-564 <RES>  
A:Cross-references: GB:S65038; NID:G410635; PID:G410636  
C:Genetics:  
A:Gene: gli  
C:Superfamily: gli transforming protein

Query Match 27.1%; Score 52; DB 2; Length 564;  
Best Local Similarity 60.0%; Pred. No. 27;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 YEVHPYGTLPVGPDP 15  
|||:|||||:  
Db 472 YEARGPSULPGLGPP 486  
|||:|||||:

## RESULT 6

C32575  
C-ski protein FB27 - chicken  
C:Species: Gallus gallus (chicken)  
C>Date: 12-Oct-1989 #sequence\_revision 12-Oct-1989 #text\_change 18-Jun-1993  
C:Accession: C32575  
R:Sutcliffe, P.; Hughes, S.H.  
Mol. Cell. Biol. 9, 4046-4051, 1989  
A:Title: Isolation and characterization of three distinct cDNAs for the chicken c-ski ge  
A:Reference number: A32575; MUID:89384635; PMID:2779576  
A:Accession: C32575  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-649 <SUT>  
A:Cross-references: GB:M28517  
C:Superfamily: ski transforming protein

Query Match 27.1%; Score 52; DB 2; Length 649;  
Best Local Similarity 35.3%; Pred. No. 31;  
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

QY 10 PVGPGPEF-----KTLRVQNLGCVVWSG 33  
|||:|||||:  
Db 67 PVMPGPFMPDSRSTERCETILGETISCFVVG 100  
|||:|||||:

## RESULT 7

B32575  
C-ski protein FB28 - chicken  
C:Species: Gallus gallus (chicken)  
C>Date: 12-Oct-1989 #sequence\_revision 12-Oct-1989 #text\_change 18-Jun-1993  
C:Accession: B32575  
R:Sutcliffe, P.; Hughes, S.H.  
Mol. Cell. Biol. 9, 4046-4051, 1989  
A:Title: Isolation and characterization of three distinct cDNAs for the chicken c-ski ge  
A:Reference number: A32575; MUID:89384635; PMID:2779576  
A:Accession: B32575  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-713 <SUT>  
A:Cross-references: GB:M28517  
C:Superfamily: ski transforming protein

Query Match 27.1%; Score 52; DB 2; Length 713;  
Best Local Similarity 35.3%; Pred. No. 35;  
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

QY 10 PVGPGPEF-----KTLRVQNLGCVVWSG 33  
|||:|||||:  
Db 67 PVMPGPFMPDSRSTERCETILGETISCFVVG 100  
|||:|||||:

## RESULT 8

TVHUSK  
transforming protein ski - human  
C:Species: Homo sapiens (man)  
C>Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 09-Jul-2004  
C:Accession: S06053

A; Introns: 59/3; 98/3; 175/2; 208/3; 372/3

Query Match 26.8%; Score 51.5; DB 2; Length 390;  
Best Local Similarity 54.5%; Pred. No. 21;  
Matches 12; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

Qy 11 VGPGPPEKTTLRVONLGCYVVS 32  
||||| : ||||| :  
Db 4 VGPGPPEKTTLRVONLGCYVVS 24

RESULT 11

S22973  
cytochrome P450 52A7 - yeast (Candida tropicalis)  
N; Alternate names: cytochrome P450alk4  
N; Contains: oxidoreductase (EC 1.-.-.-)  
C; Species: Candida tropicalis  
C; Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
R; Accession: S22973  
R; Seghezzi, W.; Meili, C.; Ruffiner, R.; Kuenzi, R.; Sanglard, D.; Fiechter, A.;  
submitted to the EMBL Data Library, June 1992  
A; Description: Isolation and characterization of additional members of the cyto  
A; Reference number: S22972  
A; Accession: S22973  
A; Molecule type: DNA  
A; Residues: 1-507 <SEG>  
A; Cross-references: UNIPROT:P30609; EMBL:Z13011; NID:G2661; PIDN:CAA78355.1; PI  
C; Genetics:  
A; Gene: CYP52A7  
C; Superfamily: Candida cytochrome P450 52A3; cytochrome P450 homology  
C; Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; oxid  
F; 303-478/Domain: cytochrome P450 homology <CYP>  
F; 456/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 26.8%; Score 51.5; DB 1; Length 507;  
Best Local Similarity 40.7%; Pred. No. 28;  
Matches 11; Conservative 6; Mismatches 9; Indels 1; Gaps 1;

Qy 8 TLPVPGPPEKTTLRVONLGCYV-VSG 33  
||||| : ||||| : ||||| :  
Db 388 TLPVGGPDGKDPILVRKMSCSIFISG 414

RESULT 12

T37694  
hypothetical protein SPAC14C4.09 - fission yeast (Schizosaccharomyces pombe)  
C; Species: Schizosaccharomyces pombe  
C; Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C; Accession: T37694  
R; Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, August 1997  
A; Reference number: Z21737  
A; Accession: T37694  
A; Status: preliminary; translated from GB/EMBL/DBJ  
A; Molecule type: DNA  
A; Residues: 1-402 <DEV>  
A; Cross-references: UNIPROT:O13716; EMBL:Z98596; PIDN:CAB11202.1; GSPDB:GNO0066  
A; Experimental source: strain 972h-; cosmid c14C4  
C; Genetics:  
A; Gene: SPDB:SPAC14C4.09  
A; Map position: 1  
C; Superfamily: Schizosaccharomyces pombe hypothetical protein SPBC646.06c

Query Match 26.6%; Score 51; DB 2; Length 402;  
Best Local Similarity 43.5%; Pred. No. 26;  
Matches 10; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 7 GTLPVPGPPEKTTLRVONLGCY 29  
||||| : ||||| : ||||| :  
Db 374 GTLASGTGPEIVDSLSYFNAY 396

RESULT 13

## A48753

NFAT transcription factor phosphoprotein - mouse (fragment)  
N:Alternate names: nuclear factor of activated T cells, phosphoprotein chain  
C:Species: Mus musculus (house mouse)  
C>Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 17-Mar-1999  
C:Accession: A48753  
R:McCaifrey, P.G.; Luo, C.; Kerppola, T.K.; Jain, J.; Badalian, T.M.; Ho, A.M.; Burgeon, S.; Science 262, 750-754, 1993  
A:Title: Isolation of the cyclosporin-sensitive T cell transcription factor NFATp.  
A:Reference number: A48753; MUID:94053710; PMID:8235597  
A:Accession: A48753  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-890 <MCC>  
A:Cross-references: GB:U02079  
C:Keywords: phosphoprotein; transcription factor

Query Match 26.6%; Score 51; DB 2; Length 890;  
Best Local Similarity 75.0%; Pred. No. 61;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 HPVGTLPVGPQP 15

DB 805 HPLGTCFVLPGP 816

## RESULT 14

S34172  
sucrose-phosphate synthase (EC 2.4.1.14) - potato  
C:Species: Solanum tuberosum (potato)  
C>Date: 22-Nov-1993 #sequence\_revision 03-Aug-1995 #text\_change 09-Jul-2004  
C:Accession: S34172  
R:Sonnewald, U.; Basner, A.  
submitted to the EMBL Data Library, June 1993  
A:Reference number: S34172  
A:Accession: S34172  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1053 <SON>  
A:Cross-references: UNIPROT:Q43845; EMBL:X73477; NID:G313264; PIDN:CAA51872.1; PID:G313264  
C:Function:  
A:Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fructose  
A:Pathway: sucrose biosynthesis  
C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology  
C:Keywords: Glycosyltransferase; hexosyltransferase; sucrose biosynthesis  
F:168-651/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 26.6%; Score 51; DB 2; Length 1053;  
Best Local Similarity 35.5%; Pred. No. 74;  
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEHPYGTLPVGPPEPKTTLRVONLGYVYV 31

DB 911 FKVCKPGTVP--PSKELKVMRIQALRCHAV 939

## RESULT 15

T03464  
probable methylenetetrahydrofolate dehydrogenase (NADP) (EC 1.5.1.5) / / methenyltetrahydrofolate dehydrogenase  
C:Species: Rhodobacter capsulatus  
C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 09-Jul-2004  
C:Accession: T03464  
R:Vlcek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.  
Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997  
A:Title: Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SB1003  
A:Reference number: 214955; MUID:97404404; PMID:9256491  
A:Accession: T03464  
A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 1-300 <VLC>  
A:Cross-references: UNIPROT:O68031; EMBL:AF010496; NID:G3128256; PIDN:AAC16117.1; PID:G3128256  
C:Genetics:  
A:Map position: 1

C:Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate dehydrogenase; hydrolase; multifunctional enzyme; oxidoreductase  
C:Keywords: hydrolase; multifunctional enzyme; oxidoreductase  
F:9-287/Domain: methylenetetrahydrofolate dehydrogenase (NAD+) homology <MTFD>  
Query Match 26.3%; Score 50.5; DB 2; Length 300;  
Best Local Similarity 38.5%; Pred. No. 22;  
Matches 10; Conservative 5; Mismatches 8; Indels 3; Gaps 1;

QY 2 EVHPYGTLPVGPPEPKTTLRVQ 24

DB 32 EHPVGLVLSISIGSPSEIEVYVRNQ 57

Search completed: April 6, 2005, 12:17:01  
Job time : 3.21966 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 4.42639 Seconds  
(without alignments)  
3058.161 Million cell updates/sec

Title: us-09-647-544-2\_COPY\_952\_986

Perfect score: 192

Sequence: 1 YEVHPYCTLPVGPGRKTLRVQNLGCYVVSGLI 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	192	100.0	1049	5	ADR41424 Human CD-
2	192	100.0	1152	4	AAB64657 Human sec
3	192	100.0	1152	4	AAB64658 Human sec
4	192	100.0	1167	3	AAY32242 Human int
5	192	100.0	1167	4	AAB64584 Human sec
6	192	100.0	1167	6	ABP99490 Human sec
7	192	100.0	1167	6	ABR00964 Human gen
8	192	100.0	1167	6	ADA44026 Human sec
9	192	100.0	1167	8	ADQ19290 Human sof
10	142	74.0	1132	3	AAY32243 Human int
11	69	35.9	1188	4	AAU14467 Human nov
12	69	35.9	1188	4	AAU14231 Human nov
13	69	35.9	1188	4	AAB50085 Human A25
14	69	35.9	1188	5	AAU10551 Human A25
15	69	35.9	1188	7	ADE09956 Novel pro
16	69	35.9	1189	8	ADH80785 Human pol
17	59	30.7	1034	3	AAB25590 Protein e
18	59	30.7	1034	6	ADA27062 Human nov
19	59	30.7	1034	8	ADE86592 Novel hum
20	59	30.7	1058	5	ADR41496 Human CD-
21	59	30.7	1120	6	ABR58365 Human NOV
22	59	30.7	1188	4	AAB30929 Amino aci
23	59	30.7	1189	3	AAB25582 ITGAl1 pr
24	59	30.7	1189	4	ABG12949 Novel hum
25	59	30.7	1189	6	ABR58364 Human NOV

26	59	30.7	1189	6	ADA27054	Ada27054 Human nov
27	59	30.7	1189	7	ADE63570	Ade63570 Human Pro
28	59	30.7	1189	8	ADE86584	Ade86584 Novel hum
29	59	30.7	1189	8	ADQ19968	Adq19968 Human sof
30	56	29.2	266	7	ABM85385	Abm85385 Human pro
31	56	29.2	347	2	AAU11069	Aau11069 12D3 anti
32	56	29.2	545	5	ABB72288	Abb72288 Murine pr
33	56	29.2	688	5	ABB72300	Abb72300 Rat prote
34	56	29.2	696	5	ABB72289	Abb72289 Rat prote
35	56	29.2	1188	4	AAB50087	Aab50087 Murine A2
36	56	29.2	1188	5	AAU10552	Aau10552 Human sec
37	55.5	28.9	987	8	ADP29492	Adp29492 Human sec
38	55	28.6	210	4	AAU62597	Aau62597 Propionib
39	55	28.6	210	6	ABM59116	Abm59116 Propionib
40	54	28.1	257	4	ABG20619	Abg20619 Novel hum
41	54	28.1	391	4	ABG20618	Abg20618 Novel hum
42	54	28.1	2139	8	ADO17150	Ado17150 Novel exp
43	53.5	27.9	53	6	ABM64899	Abm64899 Propionib
44	53.5	27.9	349	4	ABG14393	Abg14393 Novel hum
45	53	27.6	408	6	ABU39798	Abu39798 Protein e

## ALIGNMENTS

## RESULT 1

ADR41424  
ID ADR41424 standard; protein; 1049 AA.  
XX  
AC ADR41424;  
XX  
DT 07-OCT-2004 (first entry)  
XX  
DE Human CD-like molecule HAOAD02, SEQ ID NO:223.  
XX

Human; CD-like molecule; cluster of differentiation; diagnosis; prevention; immune disorder; immunodeficiency; autoimmune disorder; blood-related disorder; haematological disorder; haemostatic disorder; thrombolytic disorder; hyperproliferative disorder; cancer; tumour; apoptotic disorder; cardiovascular disorder; respiratory disorder; endocrine disorder; neovascularisation; neurological disorder; endocrine disorder; reproductive system disorder; infectious disease; gastrointestinal disorder; drug screening; tissue regeneration; chemotaxis; gene therapy; antibody therapy; drug targeting; chromosome mapping; forensic analysis; immunophenotyping; cytostatic; haemostatic; tranquiliser; vulnery; antiinflammatory; nephrotropic; cardiant; antiallergic; anti-HIV; antirheumatic; antiarthritic; antipsoriatic; immunosuppressive; vasotropic; nootropic; neuroprotective; antithyroid; thyromimetic; gynaecological; virucide; hepatotropic; antibacterial; dermatological; chromosome lq21.

Homo sapiens.

WO200226930-A2.

04-APR-2002.

25-SEP-2001; 2001WO-US029838.

26-SEP-2000; 2000US-0235484P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Birse CE;

WPI; 2002-405050/43.

N-PSDB; ADR41248.

Novel polynucleotides and polypeptides useful for treating, preventing or ameliorating cardiovascular, renal, neurovascular, and autoimmune disorders.

Claim 11; SEQ ID NO 223; 1243pp; English.

XX  
CC The invention relates to 167 novel human CD (cluster of differentiation)-  
CC like molecules (ADR41388-ADR41563) and to cDNAs encoding them (seqid.11)-  
XX  
SQ Sequence 1049 AA;  
Query Match 100.0%; Score 192; DB 5; Length 1049;  
Best Local Similarity 100.0%; Pred. No. 1.2e-17;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 YEHPYGTLPVGPDPFKTTLRVQNLGCVVVSGLI 35  
Db 834 YEHPYGTLPVGPDPFKTTLRVQNLGCVVVSGLI 868  
RESULT 2  
AAB64657  
ID AAB64657 standard; protein; 1152 AA.  
AC AAB64657;  
DT 22-MAR-2001 (first entry)  
XX Human secreted protein BLAST search protein SEQ ID NO: 167.  
DE  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
XX vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;  
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; human; secreted protein.  
OS Homo sapiens.  
XX  
XX WO2000077197-A1.  
XX 21-DEC-2000.  
XX 01-JUN-2000; 2000WO-US014934.  
XX 11-JUN-1999; 99US-0138599P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX (ROSE/) ROSEN C A.  
XX Rosen CA, Ruben SM, Komatsoulis GA;  
XX WPI; 2001-032312/04.  
XX Isolated nucleic acid molecule encoding a human secreted protein is used  
XX in preventing, treating or ameliorating a medical condition.  
XX Disclosure; Page 543-546; 558pp; English.  
XX The invention relates to the isolation of genes AAF32757-F32803 encoding  
XX the human secreted proteins AAB64549-B64594. The sequence is a search  
XX result from a BLASTX homology search. The genes and proteins are useful  
XX for preventing, ameliorating or treating medical conditions, e.g. by  
XX protein or gene therapy. The genes are isolated from a range of human  
XX tissues disclosed in the specification. The nucleic acids, proteins,  
XX antibodies and (ant)agonists are useful in the diagnosis, treatment and  
XX prevention of: (a) cancer, e.g. breast and ovarian cancer, and other  
XX cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal  
XX tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's  
XX disease, allergies, autoimmune haemolytic anaemia, autoimmune  
XX thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,  
XX rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders  
XX such as myocardial ischaemias; (d) wound healing; (e) neurological  
XX diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases  
XX such as viral, bacterial, fungal and parasitic infections  
SQ Sequence 1152 AA;  
Query Match 100.0%; Score 192; DB 4; Length 1152;  
Best Local Similarity 100.0%; Pred. No. 1.3e-17;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 YEHPYGTLPVGPDPFKTTLRVQNLGCVVVSGLI 35  
Db 952 YEHPYGTLPVGPDPFKTTLRVQNLGCVVVSGLI 986

Best Local Similarity 100.0%; Pred. No. 1.3e-17;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 YEHPYGTLPVGPDPFKTTLRVQNLGCVVVSGLI 35  
Db 952 YEHPYGTLPVGPDPFKTTLRVQNLGCVVVSGLI 986  
RESULT 3  
AAB64658  
ID AAB64658 standard; protein; 1152 AA.  
AC AAB64658;  
DT 22-MAR-2001 (first entry)  
XX Human secreted protein BLAST search protein SEQ ID NO: 168.  
DE  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
XX vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;  
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; human; secreted protein.  
OS Homo sapiens.  
XX  
XX WO2000077197-A1.  
XX 21-DEC-2000.  
XX 01-JUN-2000; 2000WO-US014934.  
XX 11-JUN-1999; 99US-0138599P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX (ROSE/) ROSEN C A.  
XX Rosen CA, Ruben SM, Komatsoulis GA;  
XX WPI; 2001-032312/04.  
XX Isolated nucleic acid molecule encoding a human secreted protein is used  
XX in preventing, treating or ameliorating a medical condition.  
XX Disclosure; Page 547-551; 558pp; English.  
XX The invention relates to the isolation of genes AAF32757-F32803 encoding  
XX the human secreted proteins AAB64549-B64594. The sequence is used as a  
XX query sequence for doing BLASTX searches to identify homologous  
XX sequences. The genes and proteins are useful for preventing, ameliorating  
XX or treating medical conditions, e.g. by protein or gene therapy. The  
XX genes are isolated from a range of human tissues disclosed in the  
XX specification. The nucleic acids, proteins, antibodies and (ant)agonists  
XX are useful in the diagnosis, treatment and prevention of: (a) cancer,  
XX e.g. breast and ovarian cancer, and other cancers of the adrenal gland,  
XX bone, bone marrow, breast, gastrointestinal tract, liver, lung, or  
XX urogenital; (b) immune disorders e.g. Addison's disease, allergies,  
XX autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
XX Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)  
XX wound healing; (e) neurological diseases e.g. cerebral anoxia and  
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
XX and parasitic infections  
SQ Sequence 1152 AA;  
Query Match 100.0%; Score 192; DB 4; Length 1152;  
Best Local Similarity 100.0%; Pred. No. 1.3e-17;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 YEHPYGTLPVGPDPFKTTLRVQNLGCVVVSGLI 35  
Db 952 YEHPYGTLPVGPDPFKTTLRVQNLGCVVVSGLI 986



RESULT 4  
 AAY32242  
 ID AAY32242 standard; protein; 1167 AA.  
 XX AC AAY32242;  
 XX DT 15-FEB-2000 (first entry)  
 XX DE Human integrin subunit alpha-10.  
 XX KW Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;  
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;  
 KW therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker.  
 XX OS Homo sapiens.  
 XX PH Key Location/Qualifiers  
 FT Peptide 1..22 /note= "signal peptide"  
 FT Protein 23..1145 /note= "mature protein"  
 FT Domain 23..1120 /note= "extracellular domain"  
 FT Modified-site 98 /note= "N-glycosylated"  
 FT Domain 162..359 /note= "I-domain"  
 FT Modified-site 336 /note= "N-glycosylated"  
 FT Modified-site 364 /note= "N-glycosylated"  
 FT Binding-site 494..502 /note= "cation binding site motif"  
 FT Binding-site 558..566 /note= "cation binding site motif"  
 FT Binding-site 620..628 /note= "cation binding site motif"  
 FT Modified-site 733 /note= "N-glycosylated"  
 FT Modified-site 839 /note= "N-glycosylated"  
 FT Modified-site 921 /note= "N-glycosylated"  
 FT Modified-site 1018 /note= "N-glycosylated"  
 FT Modified-site 1039 /note= "N-glycosylated"  
 FT Domain 1121..1145 /note= "transmembrane domain"  
 FT Domain 1122..1167 /note= "cytoplasmic domain, specifically claimed in Claim 21"  
 XX WO9951639-A1.  
 XX 14-OCT-1999.  
 XX 31-MAR-1999; 99WO-SE000544.  
 XX 02-APR-1998; 98SE-00001164.  
 XX 28-JAN-1999; 99SE-00000319.  
 XX (ACTI-) ACTIVE BIOTECH AB.  
 XX Lundgren-Akerlund E;  
 XX WPI; 2000-052639/04.  
 XX N-PSDB; AAZ34719.  
 XX New isolated integrin subunit alpha-10, used as a marker or target molecule for cells during development, regeneration and pathological

PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or  
 XX inflammation.  
 XX Claim 1; Fig 6; 90pp; English.  
 XX This sequence represents novel human chondrocyte integrin subunit alpha-  
 CC 10 (ISa10). A splice variant is given in AAY32243. The invention relates  
 CC to a recombinant or isolated integrin heterodimer comprising the alpha10  
 CC subunit in association with subunit beta (especially beta-1). The  
 CC heterodimer and the subunit alpha-10 can be used as markers or targets of  
 CC all types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts.  
 CC They can also be used: for treating pathological conditions involving  
 CC ISa10, such as damage to cartilage, trauma, rheumatoid arthritis or  
 CC osteoarthritis; for detecting the formation of cartilage during embryonal  
 CC development, physiological or therapeutic repair of cartilage, or  
 CC detecting regeneration of cartilage or chondrocytes during  
 CC transplantation of cartilage or chondrocytes; for selection and analysis  
 CC for sorting, isolating or purification of chondrocytes; and as a target for  
 CC anti-adhesive drugs or molecules in tendon, ligament, skeletal muscle or  
 CC other tissues where adhesion impairs the function of the tissue (all  
 CC claimed). ISa10 binding entities can be used to determine the  
 CC differentiation-state of cells during embryonic development, angiogenesis  
 CC or development of cancer, in pathological conditions such as rheumatoid  
 CC arthritis, osteoarthritis or cancer, in tissue regeneration or in  
 CC therapeutic and physiological repair of cartilage (claimed). A  
 CC vaccine comprising the integrin heterodimer or subunit alpha-10 is also  
 CC claimed. ISa10 polynucleotides, vectors, host cells and methods of  
 CC producing recombinant ISa10 are also claimed  
 XX SQ Sequence 1167 AA;  
 Query Match 100.0%; Score 192; DB 3; Length 1167;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-17;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 YEYHPYCTLPVGPPEPKTTLRVQNLGCVVSGLI 35  
 Db 952 YEYHPYCTLPVGPPEPKTTLRVQNLGCVVSGLI 986  
 RESULT 5  
 AAB64584  
 ID AAB64584 standard; protein; 1167 AA.  
 XX AC AAB64584;  
 XX DT 22-MAR-2001 (first entry)  
 XX Human secreted protein #37.  
 DE Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
 KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.  
 XX OS Homo sapiens.  
 XX WO200077197-A1.  
 XX 21-DEC-2000.  
 XX 01-JUN-2000; 2000WO-US014934.  
 XX 11-JUN-1999; 99US-0138599P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX (ROSE/) ROSEN C A.  
 XX Rosen CA, Ruben SM, Komatsoulis GA;  
 XX WPI; 2001-032312/04.

DR N-PSDB; AAF32793.  
 XX Isolated nucleic acid molecule encoding a human secreted protein is used  
 PT in preventing, treating or ameliorating a medical condition.  
 XX  
 PS Claim 11; Page 496-500; 558pp; English.  
 XX  
 CC Sequences AAB64549-B64594 represent the amino acid sequences of 47 human  
 CC secreted proteins encoded by the genes AAF32757-F32803. The genes and  
 CC proteins are useful for preventing, ameliorating or treating medical  
 CC conditions, e.g. by protein or gene therapy. The genes are isolated from  
 CC a range of human tissues disclosed in the specification. The nucleic  
 CC acids, proteins, antibodies and (ant)agonists are useful in the  
 CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
 CC ovarian cancer, and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b)  
 CC immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
 CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
 CC / (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
 CC infectious diseases such as viral, bacterial, fungal and parasitic  
 CC infections  
 XX  
 SQ Sequence 1167 AA;  
 Query Match 100.0%; Score 192; DB 4; Length 1167;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-17; Indels 0; Gaps 0;  
 Matches 35; Conservative 0; Mismatches 0;  
 QY 1 YEHPYGTLPVGPGEFKTTLRVQNLGCVVSGLI 35  
 DB 952 YEHPYGTLPVGPGEFKTTLRVQNLGCVVSGLI 986  
 RESULT 6  
 ABR00964  
 ID ABR00964 standard; protein; 1167 AA.  
 XX  
 AC ABR00964;  
 XX  
 DT 26-MAR-2003 (first entry)  
 XX  
 DE Human secreted protein SEQ ID NO 434.  
 XX  
 KW Human; secreted protein; neurotropic; neuroprotective; cytostatic;  
 KW virucide; dermatological; immunosuppressive; antineoplastic; anti-HIV;  
 KW vulnary; antibacterial; antiparkinsonian; antineoplastic; antianaemic;  
 KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;  
 KW antiinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;  
 KW antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;  
 KW cardiovascular disorder; neurological disease; nephrotropic;  
 KW gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200277186-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 26-MAR-2002; 2002WO-US009198.  
 XX  
 PR 27-MAR-2001; 2001US-0278650P.  
 PR 12-SEP-2001; 2001US-00950082.  
 PR 12-SEP-2001; 2001US-00950083.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI; 2003-040583/03.  
 DR N-PSDB; AAB64594.  
 XX  
 PT New human secreted proteins and nucleic acids, useful for detecting or  
 PT treating cancer or other hyperproliferative disorders, autoimmune  
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.

PT New human secreted proteins encoded by genes contained in cDNA clones  
 PT (e.g. HGAC19), useful for preventing, treating or diagnosing e.g. AIDS,  
 PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or  
 PT West Nile fever.  
 XX  
 PS Claim 1; Page 1395-1398; 2423pp; English.  
 XX  
 CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the  
 CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,  
 CC treating or ameliorating medical conditions e.g. by protein or gene  
 CC therapy. The genes are isolated from a range of human tissues disclosed  
 CC in the specification. The nucleic acids, proteins, antibodies and  
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:  
 CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the  
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
 CC lung or urogenital; (b) immune disorders e.g. Addison's disease,  
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as  
 CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.  
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,  
 CC bacterial, fungal and parasitic infections  
 XX  
 SQ Sequence 1167 AA;  
 Query Match 100.0%; Score 192; DB 6; Length 1167;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-17; Indels 0; Gaps 0;  
 Matches 35; Conservative 0; Mismatches 0;  
 QY 1 YEHPYGTLPVGPGEFKTTLRVQNLGCVVSGLI 35  
 DB 952 YEHPYGTLPVGPGEFKTTLRVQNLGCVVSGLI 986  
 RESULT 7  
 ABR00964  
 ID ABR00964 standard; protein; 1167 AA.  
 XX  
 AC ABR00964;  
 XX  
 DT 12-MAY-2003 (first entry)  
 XX  
 DE Human gene 18-encoded secreted protein HAOAG15, SEQ ID NO:445.  
 XX  
 KW Human; secreted protein; cancer; tumour; hyperproliferative disorder;  
 KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;  
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;  
 KW drug screening; chromosome identification; chromosome mapping;  
 KW cytostatic; gene therapy; antineoplastic; immunomodulator; anti-HIV;  
 KW antianaemic; vulnary; chromosome 1q21.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200277013-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 26-MAR-2002; 2002WO-US009370.  
 XX  
 PR 27-MAR-2001; 2001US-0278650P.  
 PR 12-SEP-2001; 2001US-00950082.  
 PR 12-SEP-2001; 2001US-00950083.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI; 2003-040578/03.  
 DR N-PSDB; ABZ73298.  
 XX  
 PT New human secreted proteins and nucleic acids, useful for detecting or  
 PT treating cancer or other hyperproliferative disorders, autoimmune  
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.

XX Claim 13; Page 1387-1390; 2474pp; English.

XX AB273281-AB273697 represent cDNAs corresponding to 391 human secreted

CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.

CC AB273698-AB274687 represent human secreted protein genomic fragments. The

CC invention also encompasses antibodies specific for the secreted proteins,

CC the use of the secreted proteins in drug screening and recombinant

CC vectors and host cells comprising a nucleic acid of the invention. The

CC secreted proteins are thought to be involved in biological activities

CC associated with cellular signalling, cellular differentiation, cell

CC migration, prohormone activation and neurotransmitter activity. The

CC secreted proteins, nucleic acids encoding them, antibodies or antibody

CC fragments specific for the secreted proteins, and modulators of protein

CC activity are useful for diagnosing or treating cancers or other

CC hyperproliferative disorders. Additionally, the secreted proteins and

CC their nucleic acids may also be used in the treatment of autoimmune

CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS

CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote

CC wound healing. Nucleic acids of the invention may be used for chromosome

CC identification, chromosome mapping, in gene therapy, for identifying

CC individuals from minute biological samples, as hybridisation probes, and

CC as molecular weight markers. The present sequence represents a human

CC secreted protein of the invention

XX

XX Sequence 1167 AA;

Query Match 100.0%; Score 192; DB 6; Length 1167;

Best Local Similarity 100.0%; Pred. No. 1.4e-17;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEHPYGTLPVGPGEFKTTLRVQNLGCVVSGLI 35

Db 952 YEHPYGTLPVGPGEFKTTLRVQNLGCVVSGLI 986

RESULT 8

ADA44026

ID ADA44026 standard; protein; 1167 AA.

XX

XX ADA44026;

XX

XX 20-NOV-2003 (first entry)

XX

XX Human secreted protein SEQ ID 214.

XX

XX Gene therapy; human; Antidiabetic; Anorectic; Ophthalmological;

XX Neuroprotective; Cerebroprotective; Antianemic.

XX

XX Homo sapiens.

XX

XX WO2003000865-A2.

XX

XX 03-JAN-2003.

XX

XX 26-MAR-2002; 2002WO-US009105.

XX

XX 27-MAR-2001; 2001US-0278650P.

XX

XX 12-SEP-2001; 2001US-00950082.

XX

XX 12-SEP-2001; 2001US-00950083.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

XX

XX Rosen CA, Ruben SM;

XX

XX WPI; 2003-184045/18.

XX

XX N-PSDB; ADA43832.

XX

XX A human secreted protein and nucleic acids useful for preparing a

XX diagnostic or pharmaceutical composition for diagnosing or treating

XX diabetes or conditions related to diabetes, e.g. hyperglycemia, obesity,

XX retinopathy, neuropathy.

XX

Claim 1; SEQ ID NO 214; 701pp; English.

XX

XX The invention relates to novel genes and their fragments which are useful

CC for preventing, treating or ameliorating medical conditions e.g. by

CC protein or gene therapy. The genes are isolated from a range of human

CC tissues disclosed in the specification. The nucleic acids and proteins

CC are useful in the diagnosis, treatment and prevention of conditions

CC related to diabetes, e.g. hyperglycaemia, obesity, retinopathy,

CC polynuropathy, atherosclerosis, anaemia, stroke, gangrene, impotence,

CC infection, cataract, renal disorders, or endocrine disorders. The present

CC sequence was used to illustrate the invention.

XX

XX Sequence 1167 AA;

Query Match 100.0%; Score 192; DB 6; Length 1167;

Best Local Similarity 100.0%; Pred. No. 1.4e-17;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEHPYGTLPVGPGEFKTTLRVQNLGCVVSGLI 35

Db 952 YEHPYGTLPVGPGEFKTTLRVQNLGCVVSGLI 986

RESULT 9

ADQ19290

ID ADQ19290 standard; protein; 1167 AA.

XX

XX ADQ19290;

XX

XX 26-AUG-2004 (first entry)

XX

XX Human soft tissue sarcoma-upregulated protein - SEQ ID 2109.

XX

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.

XX

XX Homo sapiens.

XX

XX WO2004048938-A2.

XX

XX 10-JUN-2004.

XX

XX 26-NOV-2003; 2003WO-US038193.

XX

XX 26-NOV-2002; 2002US-0429739P.

XX

XX (PROT-) PROTEIN DESIGN LABS INC.

XX

XX Aziz N, Ginsburg WM, Zlotnik A;

XX

XX WPI; 2004-441208/41.

XX

XX Early detection of soft tissue sarcoma comprises determining expression

XX of a gene in a first soft tissue sample and a normal soft tissue sample

XX and comparing the gene expression, also useful in treating soft tissue

XX sarcoma.

XX

XX Example 2; SEQ ID NO 2109; 210pp; English.

XX

XX The invention relates to a novel method for detecting soft tissue sarcoma

CC which comprises obtaining a first soft tissue sample from an individual

CC and a normal soft tissue sample from the same or different individual,

CC determining the expression of a gene in both samples and comparing the

CC expression of the gene in both soft tissue samples, where a higher level

CC of protein expression in the first soft tissue sample indicates the

CC presence of soft tissue sarcoma. The method of the invention has

CC cytostatic applications and may be useful for detecting soft tissue

CC sarcoma, possibly via gene therapy or vaccine production. The nucleic

CC acid sequences may be useful in diagnostic and screening applications.

CC The current sequence is that of a human soft tissue sarcoma-upregulated

CC protein of the invention. The current sequence is not shown within the

CC specification per se but was submitted in CD format by the inventor.

XX

XX Sequence 1167 AA;

Query Match 100.0%; Score 192; DB 8; Length 1167;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-17;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEHPYGTLPVGPPEFKTTLRVQNLGCVVSGLI 35  
 |||||  
 Db 952 YEHPYGTLPVGPPEFKTTLRVQNLGCVVSGLI 986

## RESULT 10

AA132243  
 ID AA132243 standard; protein; 1132 AA.

XX

AC AA132243;

DT 15-FEB-2000 (first entry)

XX Human integrin subunit alpha-10 splice variant.

XX Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;  
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;  
 KW therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker;  
 KW splice variant.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..22 "signal peptide"

FT Protein 23..1132

FT "mature protein"

XX W09951639-A1.

XX 14-OCT-1999.

XX 31-MAR-1999; 99WO-SE000544.

XX 02-APR-1998; 98SE-00001164.

XX 28-JAN-1999; 99SE-00000319.

XX (ACT1-) ACTIVE BIOTECH AB.

XX Lundgren-Akerlund E;

XX WPI; 2000-052639/04.

XX N-PSDB; AA234720.

XX New isolated integrin subunit alpha-10, used as a marker or target  
 molecule for cells during development, regeneration and pathological  
 conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or  
 inflammation.

XX Claim 1; Page 43-48; 90pp; English.

XX This sequence represents a splice variant of novel human chondrocyte  
 integrin subunit alpha-10 (ISa10). It is identical to ISa10 (see  
 AA132242) except for deletion of amino acids 975-986. The invention  
 relates to a recombinant or isolated integrin heterodimer comprising the  
 alpha10 subunit in association with subunit beta (especially beta-1). The  
 heterodimer, subunit alpha-10 or splice variant can be used as a marker  
 or target of all types of cells, e.g. of chondrocytes, osteoblasts and  
 fibroblasts. They can also be used for treating pathological conditions  
 involving ISa10, such as damage to cartilage, trauma, rheumatoid  
 arthritis or osteoarthritis; for detecting the formation of cartilage  
 during embryonal development, physiological or therapeutic repair of  
 cartilage, or detecting regeneration of cartilage or chondrocytes during  
 transplantation of cartilage or chondrocytes; for selection and analysis  
 or for sorting, isolating or purification of chondrocytes and for in  
 vitro studies of differentiation of chondrocytes; and as a target for  
 anti-adhesive drugs or molecules in tendon, ligament, skeletal muscle or  
 other tissues where adhesion impairs the function of the tissue (all

CC claimed). ISa10 binding entities can be used to determine the  
 CC differentiation-state of cells during embryonic development, angiogenesis  
 CC or development of cancer, in pathological conditions such as rheumatoid  
 CC arthritis, osteoarthritis or cancer, in tissue regeneration or in  
 CC therapeutic and physiological repair of cartilage (claimed). A  
 CC vaccine comprising the integrin heterodimer or subunit alpha-10 is also  
 CC claimed. ISa10 polynucleotides, vectors, host cells and methods of  
 CC producing recombinant ISa10 are also claimed

XX Sequence 1132 AA;

Query Match 74.0%; Score 142; DB 3; Length 1132;  
 Best Local Similarity 83.3%; Pred. No. 1.2e-10;  
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 YEHPYGTLPVGPPEFKTTLRVQNLGCVV 30

Db 952 YEHPYGTLPVGPPEFKTTLRTNNASCIV 981

## RESULT 11

AAU14467  
 ID AAU14467 standard; protein; 1188 AA.

XX AAU14467;

XX 24-OCT-2001 (first entry)

XX Human novel protein #338.

XX Human; novel protein; Antianaemic; osteopathic; antiinflammatory;  
 KW immunomodulatory; cytostatic; neuroprotective; vulnary; nootropic;  
 KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;  
 KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;  
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
 KW tissue regeneration; immune disorder.

XX Homo sapiens.

XX W0200155437-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US002623.

XX 25-JAN-2000; 2000US-00491404.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-451939/48.

XX N-PSDB; AA22772.

XX Isolated polypeptides useful for treating anti-inflammatory diseases,  
 XX nervous system disorders, and for regenerating bone and cartilage.

XX Example 4; Page 828-831; 894pp; English.

XX The invention relates to polynucleotides encoding novel human proteins or  
 CC their active domains. The polypeptides, polynucleotides and antibodies  
 CC raised against the polypeptides are used in a method of treatment of a  
 CC mammal and prevention of disorders caused by the aberrant protein  
 CC expression or activity. The polypeptides can be used as molecular weight  
 CC markers, food supplements, and in antibody production. The polypeptides  
 CC are used to identify compounds which bind to the polypeptides.  
 CC Polynucleotides of the invention are used as probes and primers, for  
 CC sequencing, for chromosome or gene mapping, in the production of  
 CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene  
 CC therapy. Polypeptides of the invention can be used to target drugs to a  
 CC tumour, in assays to determine biological activity, to raise  
 CC antibodies/elicite an immune response, to determine quantitative protein

CC levels, as tissue markers, and to isolate receptors or ligands.  
 CC Polypeptides of the invention may also be useful in treating platelet  
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
 CC the proliferation, differentiation and survival of stem cells, as a  
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-  
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory  
 CC diseases, nervous system disorders, and infection. The present sequence  
 CC represents a protein of the invention  
 XX  
 SQ Sequence 1188 AA;  
 Query Match 35.9%; Score 69; DB 4; Length 1188;  
 Best Local Similarity 40.5%; Pred. No. 1.9;  
 Matches 15; Conservative 6; Mismatches 14; Indels 2; Gaps 1;  
 QY 1 YEVPYGTLP--VGPGEFKTTLRVQNLGCVVWSGLI 35  
 DB 968 YEVPKNSLERYDGI GPPFCIFRIQNLGLFPIHGM 1004  
 RESULT 12  
 AAU14231  
 ID AAU14231 standard; protein; 1188 AA.  
 AC AAU14231;  
 DT 24-OCT-2001 (first entry)  
 DE Human novel protein #102.  
 KW Human; novel protein; Antianaemic; osteopathic; anti-inflammatory;  
 KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;  
 KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;  
 KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;  
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
 KW tissue regeneration; immune disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200155437-A2.  
 XX  
 XX 02-AUG-2001.  
 XX  
 XX 25-JAN-2001; 2001WO-US002623.  
 XX  
 XX 25-JAN-2000; 2000US-00491404.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Drmanac RT;  
 XX  
 XX WPI; 2001-451939/48.  
 XX  
 XX N-PSDB; AAS22536.  
 XX  
 XX Isolated polypeptides useful for treating anti-inflammatory diseases,  
 XX nervous system disorders, and for regenerating bone and cartilage.  
 XX  
 XX Example 4; Page 578-581; 894pp; English.  
 XX  
 XX The invention relates to polynucleotides encoding novel human proteins or  
 XX their active domains. The polypeptides, polynucleotides and antibodies  
 XX raised against the polypeptides are used in a method of treatment of a  
 XX mammal and prevention of disorders caused by the aberrant protein  
 XX expression or activity. The polypeptides can be used as molecular weight  
 XX markers, food supplements, and in antibody production. The polypeptides  
 XX are used to identify compounds which bind to the polypeptides.  
 XX Polynucleotides of the invention are used as probes and primers, for  
 XX sequencing, for chromosome or gene mapping, in the production of

CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene  
 CC therapy. Polypeptides of the invention can be used to target drugs to a  
 CC tumour, in assays to determine biological activity, to raise  
 CC antibodies/elicite an immune response, to determine quantitative protein  
 CC levels, as tissue markers, and to isolate receptors or ligands.  
 CC Polypeptides of the invention may also be useful in treating platelet  
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
 CC the proliferation, differentiation and survival of stem cells, as a  
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-  
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory  
 CC diseases, nervous system disorders, and infection. The present sequence  
 CC represents a protein of the invention  
 XX  
 SQ Sequence 1188 AA;  
 Query Match 35.9%; Score 69; DB 4; Length 1188;  
 Best Local Similarity 40.5%; Pred. No. 1.9;  
 Matches 15; Conservative 6; Mismatches 14; Indels 2; Gaps 1;  
 QY 1 YEVPYGTLP--VGPGEFKTTLRVQNLGCVVWSGLI 35  
 DB 968 YEVPKNSLERYDGI GPPFCIFRIQNLGLFPIHGM 1004  
 RESULT 13  
 AAB50085  
 ID AAB50085 standard; protein; 1188 AA.  
 AC AAB50085;  
 DT 19-MAR-2001 (first entry)  
 DE Human A259.  
 KW Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;  
 KW kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;  
 KW rheumatoid arthritis.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 XX Domain 1..1141  
 XX /label= Extracellular\_domain  
 XX Peptide 1..22  
 XX /label= Signal\_peptide  
 XX Protein 23..1188  
 XX /label= Mature\_protein  
 XX Domain 39..74  
 XX /label= Integrin\_alphasubunit\_repeat\_domain\_#1  
 XX Domain 115..157  
 XX /label= Integrin\_alphasubunit\_repeat\_domain\_#2  
 XX Domain 164..345  
 XX /label= I\_domain  
 XX Domain 367..392  
 XX /label= Integrin\_alphasubunit\_repeat\_domain\_#3  
 XX Domain 421..455  
 XX /label= Integrin\_alphasubunit\_repeat\_domain\_#4  
 XX Domain 478..516  
 XX /label= Integrin\_alphasubunit\_repeat\_domain\_#5  
 XX Domain 540..575  
 XX /label= Integrin\_alphasubunit\_repeat\_domain\_#6  
 XX Domain 602..640  
 XX /label= Integrin\_alphasubunit\_repeat\_domain\_#7  
 XX Domain 1142..1164  
 XX /label= Transmembrane\_domain  
 XX Domain 1165..1188  
 XX /label= Cytoplasmic\_domain  
 XX  
 XX WO200073339-A1.



**RESULT 15**

AD09956	AD09956 standard; protein; 1188 AA.
ID	AD09956 standard; protein; 1188 AA.
XX	
AC	AD09956;
XX	
DT	29-JAN-2004 (first entry)
XX	
DE	Novel protein-related contig polypeptide sequence #544.
XX	
KW	novel gene; novel protein; tissue marker; molecular weight marker;
KW	chromosome marker; genetic disorder; contig.
XX	
OS	Unidentified.
XX	
PN	W02003054152-A2.
XX	
PD	03-JUL-2003.
XX	
PF	10-DEC-2002; 2002W0-US039555.
XX	
PR	10-DEC-2001; 2001US-0339739P.
PR	11-DEC-2001; 2001US-0339453P.
PR	14-MAR-2002; 2002US-0365091P.
PR	14-MAR-2002; 2002US-0365384P.
PR	12-APR-2002; 2002US-0372381P.
PR	12-APR-2002; 2002US-0372615P.
PR	22-APR-2002; 2002US-00128558.
PR	24-APR-2002; 2002US-0376045P.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Aeundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
PI	Chosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
PI	Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
XX	
DR	WPI; 2003-569235/53.
XX	
PT	New polynucleotides, useful for expressing recombinant proteins for
PT	analysis, characterization or therapeutic use, or as markers for tissues
PT	in which the corresponding protein is preferentially expressed.
XX	
PS	Disclosure; SEQ ID NO 3022; 1177pp; English.
XX	
CC	The invention comprises the amino acid and coding sequences of novel
CC	proteins. The DNA and protein sequences of the invention are useful as:
CC	markers for tissues in which the corresponding protein is preferentially
CC	expressed; as molecular weight markers on gels; as chromosome markers or
CC	tags; to identify chromosomes or to map related gene positions; and to
CC	compare with endogenous DNA sequences in patients to identify potential
CC	genetic disorders. The present amino acid sequence was used in the
CC	exemplification of the invention.
XX	
SQ	Sequence 1188 AA;
XX	
Query Match	35.9%; Score 69; DB 7; Length 1188;
Best Local Similarity	40.5%; Pred. NO. 1.9;
Matches 15; Conservative	6; Mismatches 14; Indels 2; Gaps

Search completed: April 6, 2005, 12:22:28  
Job time : 5.42639 secs

Page Blank (uspto)



Result No.	Score	Query Match	Length	DB	ID	Description
1	192	100.0	1167	16	US-10-741-601-531	Sequence 531, App
2	192	100.0	1177	15	US-10-741-601-532	Sequence 532, App
3	69	35.9	1188	15	US-10-741-601-538	Sequence 336, App
4	69	35.9	1188	15	US-10-741-601-538	Sequence 810, App
5	30.7	1034	10	US-09-391-265-810	Sequence 43, Appl	
6	59	30.7	1034	10	US-09-391-265-810	Sequence 43, Appl
7	59	30.7	1034	10	US-09-384-130-43	Sequence 6, Appli
8	59	30.7	1120	15	US-10-741-601-531	Sequence 35, Appl
9	59	30.7	1189	10	US-09-384-130-35	Sequence 35, Appl
10	59	30.7	1189	15	US-09-384-130-35	Sequence 35, Appl
11	57.5	29.9	3338	14	US-10-741-601-532	Sequence 4, Appli
12	56.5	23.4	1126	16	US-10-741-601-532	Sequence 8464, Ap
13	56	29.2	266	13	US-10-437-963-182396	Sequence 182396,
					US-10-087-192-570	Sequence 570, App

```
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741.601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 532
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-741-601-532

Query Match      100.0%; Score 192; DB 16; Length 1177;
Best Local Similarity 100.0%; Pred. No. 4.1e-17;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 YEVHPYGTLP--VGPGEPEKTTLRVQNLGCVYVSGLI 35
Db 952 YEVHPYGTLPVGPGEPEKTTLRVQNLGCVYVSGLI 986

RESULT 3
US-10-291-265-338
; Sequence 338, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291.265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 338
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-338

Query Match      35.9%; Score 69; DB 15; Length 1188;
Best Local Similarity 40.5%; Pred. No. 2.1;
Matches 15; Conservative 6; Mismatches 14; Indels 2; Gaps 1;

Qy 1 YEVHPYGTLP--VGPGEPEKTTLRVQNLGCVYVSGLI 35
Db 968 YEVKPNSSLERYDGIPFPSCIPRIQNLGLFPIHGMM 1004

RESULT 4
US-10-291-265-810
; Sequence 810, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291.265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
```

```
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: 09/633.870
; CURRENT FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 810
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-810

Query Match      35.9%; Score 69; DB 15; Length 1188;
Best Local Similarity 40.5%; Pred. No. 2.1;
Matches 15; Conservative 6; Mismatches 14; Indels 2; Gaps 1;

Qy 1 YEVHPYGTLP--VGPGEPEKTTLRVQNLGCVYVSGLI 35
Db 968 YEVKPNSSLERYDGIPFPSCIPRIQNLGLFPIHGMM 1004

RESULT 5
US-09-984-130-43
; Sequence 43, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984.130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-43

Query Match      30.7%; Score 59; DB 10; Length 1034;
Best Local Similarity 41.7%; Pred. No. 42;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 12 GPGPEPKTTLRVQNLGCVYVSGLI 35
Db 981 GIGPPFPSCIPRIQNLGLFPIHGIM 1004

RESULT 6
US-09-836-353A-43
; Sequence 43, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P1
; CURRENT APPLICATION NUMBER: US/09/836.353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-353A-43

Query Match      30.7%; Score 59; DB 10; Length 1034;
Best Local Similarity 41.7%; Pred. No. 42;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Oy 12 GPGPEFKTTLRVQNLGCVVWSGLI 35
Db 981 GIGPPFSCIFRIQNLGLFPIHGIM 1004

RESULT 7
US-10-262-839-6
; Sequence 6, Application US/10262839
; Publication No. US20040038877A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook, John,
; APPLICANT: Anderson, David W.,
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Burgess, Catherine,
; APPLICANT: Catterton, Elina,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Ji, Weizhen,
; APPLICANT: Kekuda, Rameah,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Smithson, Glenda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Taupier, Raymond, Jr.,
; APPLICANT: Vernet, Corine,
; APPLICANT: Voss, Edward,
; APPLICANT: Zerhusen, Brian,
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-462A
; CURRENT APPLICATION NUMBER: US/10/262,839
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,101
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/327,342
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 60/328,044
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/374,738
; PRIOR FILING DATE: 2002-04-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367

; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 6
; LENGTH: 1120
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-262-839-6

Query Match      30.7%; Score 59; DB 15; Length 1120;
Best Local Similarity 41.7%; Pred. No. 46;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Oy 12 GPGPEFKTTLRVQNLGCVVWSGLI 35
Db 912 GIGPPFSCIFRIQNLGLFPIHGIM 935

RESULT 8
US-09-984-130-35
; Sequence 35, Application US/09984130
; Publication No. US2003005231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-35

Query Match      30.7%; Score 59; DB 10; Length 1189;
Best Local Similarity 41.7%; Pred. No. 49;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Oy 12 GPGPEFKTTLRVQNLGCVVWSGLI 35
Db 981 GIGPPFSCIFRIQNLGLFPIHGIM 1004

RESULT 9
US-09-836-353A-35
; Sequence 35, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
```

```
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-836-353A-35

Query Match      30.7%; Score 59; DB 10; Length 1189;
Best Local Similarity 41.7%; Pred. No. 49;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 12 GPGPEFKTTLRVONLGCYVVSGLI 35
Db 981 GIGPFFSCIFRIQNGLGFLPIHGIM 1004

RESULT 10
US-10-262-839-4
; Sequence 4, Application US/10262839
; Publication No. US20040038877A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John,
; APPLICANT: Anderson, David W.,
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Burgess, Catherine,
; APPLICANT: Catterton, Elina,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Ji, Weizhen,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Smithson, Glenda,
; APPLICANT: Spytsek, Kimberly,
; APPLICANT: Taupier, Raymond, Jr.,
; APPLICANT: Vernet, Corine,
; APPLICANT: Voss, Edward,
; APPLICANT: Zerkhusen, Brian,
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-462A
; CURRENT APPLICATION NUMBER: US/10/262,839
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,101
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/327,342
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 60/328,044
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,849
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/374,738
; PRIOR FILING DATE: 2002-04-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 367
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 4
; LENGTH: 1189
```

```
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-262-839-4

Query Match      30.7%; Score 59; DB 15; Length 1189;
Best Local Similarity 41.7%; Pred. No. 49;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 12 GPGPEFKTTLRVONLGCYVVSGLI 35
Db 981 GIGPFFSCIFRIQNGLGFLPIHGIM 1004

RESULT 11
US-10-156-761-8464
; Sequence 8464, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HORIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8464
; LENGTH: 3338
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8464

Query Match      29.9%; Score 57.5; DB 14; Length 3338;
Best Local Similarity 56.5%; Pred. No. 2.3e+02;
Matches 13; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 8 TLPVGPPEFKTTLRVONLGCYV 30
Db 2091 TLP-GPDPFSPALRLRGLGLQV 2112

RESULT 12
US-10-437-963-182396
; Sequence 182396, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 182396
; LENGTH: 1126
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
```

```
; NAME/KEY: unsure
; LOCATION: (1)...(1126)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_7958C.1.pap
US-10-437-963-182396

Query Match      29.2%; Score 56.5; DB 16; Length 1126;
Best Local Similarity 52.2%; Pred. No. 1e+02;
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

Qy 13 PGPEFKTTLRVQNLGCVVWSGL 34
Db 322 PGSEFSTPLKVLPLGCVFVILGM 344

RESULT 13
US-10-087-192-570
; Sequence 570, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 570
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-087-192-570

Query Match      29.2%; Score 56; DB 13; Length 266;
Best Local Similarity 47.8%; Pred. No. 25;
Matches 11; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Qy 5 PYGTLVPGPGPEFKTTLRVQNLG 27
Db 76 PRGMGMPGPGPQTDPLSLQNYG 98

RESULT 14
US-09-866-050A-500
; Sequence 500, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 500
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Mouse
US-09-866-050A-500

; NAME/KEY: unsure
; LOCATION: (1)...(1126)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_7958C.1.pap
US-10-437-963-182396

Query Match      29.2%; Score 56; DB 10; Length 545;
Best Local Similarity 41.7%; Pred. No. 54;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 12 GPGPEFKTTLRVQNLGCVVWSGLI 35
Db 338 GIGPPFNCVFKVQNLGFFPIHGM 361

RESULT 15
US-09-866-050A-624
; Sequence 624, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D. Isolated From Skin Cells
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 624
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Rat
US-09-866-050A-624

Query Match      29.2%; Score 56; DB 10; Length 688;
Best Local Similarity 41.7%; Pred. No. 69;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 12 GPGPEFKTTLRVQNLGCVVWSGLI 35
Db 481 GIGPPFNCVFKVQNLGFFPIHGM 504

Search completed: April 6, 2005, 12:53:01
Job time : 5.37157 secs
```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:52 ; Search time 1.49374 Seconds  
(without alignments)  
1749.117 Million cell updates/sec

Title: US-09-647-544-2\_COPY\_952\_986

Perfect score: 192

Sequence: 1 YEHVPYGLPVGPGPEKTLRVQNLGCVVSGLI 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database :
- 1: Issued Patents AA:\*
  - 2: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*
  - 1: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*
  - 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*
  - 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*
  - 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*
  - 6: /cgn2\_6/ptodata/1/iaa/backfiles.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	30.7	1217	4	US-09-949-016-7892
2	52.5	27.3	469	4	US-09-270-767-33880
3	52.5	27.3	469	4	US-09-270-767-49097
4	52	27.1	750	3	US-08-202-841A-2
5	51	26.6	846	1	US-08-356-354-2
6	51	26.6	846	2	US-08-778-656-2
7	51	26.6	846	4	US-09-376-045-2
8	51	26.6	890	1	US-08-145-006C-5
9	51	26.6	890	4	US-09-546-013-18
10	51	26.6	890	5	PCT-US94-00545-5
11	51	26.6	908	1	US-08-356-354-6
12	51	26.6	908	2	US-08-778-656-6
13	51	26.6	908	4	US-09-376-045-6
14	51	26.6	1053	4	US-09-394-272-6
15	51	26.6	1054	1	US-08-356-354-4
16	51	26.6	1054	2	US-08-778-656-4
17	51	26.6	1054	4	US-09-376-045-4
18	50	26.0	738	3	US-08-989-385-1
19	50	26.0	738	4	US-09-593-826-1
20	50	26.0	738	4	US-09-949-016-6020
21	50	26.0	738	4	US-09-949-016-11542
22	50	26.0	753	4	US-09-902-540-15167
23	49.5	25.8	305	4	US-09-529-239D-31
24	49	25.5	1109	4	US-09-621-976-7129
25	48	25.0	51	4	US-09-902-540-15376
26	48	25.0	138	4	US-09-248-796A-16892
27	48	25.0	211	4	US-09-248-796A-16892

28	48	25.0	217	4	US-09-252-991A-24818	Sequence 24818, A
29	48	25.0	266	4	US-09-248-796A-18478	Sequence 18478, A
30	48	25.0	668	3	US-09-697-367-2	Sequence 2, Appli
31	48	25.0	668	4	US-09-918-909A-2	Sequence 2, Appli
32	48	25.0	1261	4	US-09-248-796A-16620	Sequence 16620, A
33	48	25.0	1289	1	US-07-876-280-4	Sequence 4, Appli
34	48	25.0	1289	1	US-07-675-772-4	Sequence 4, Appli
35	48	25.0	1289	1	US-08-063-170-4	Sequence 4, Appli
36	48	25.0	1289	1	US-08-158-232-4	Sequence 4, Appli
37	48	25.0	1289	1	US-08-304-626-4	Sequence 4, Appli
38	48	25.0	1289	1	US-08-316-301A-4	Sequence 4, Appli
39	48	25.0	1289	2	US-08-611-928-4	Sequence 4, Appli
40	48	25.0	1289	3	US-09-173-891-4	Sequence 4, Appli
41	48	25.0	1289	3	US-09-076-137-4	Sequence 4, Appli
42	48	25.0	1289	4	US-09-738-363-4	Sequence 4, Appli
43	48	25.0	1289	5	PCT-US92-03624-4	Sequence 4, Appli
44	48	25.0	1289	6	5281530-3	Patent No. 5281530
45	48	25.0	1289	6	5426049-4	Patent No. 5426049

ALIGNMENTS

RESULT 1

US-09-949-016-7892  
; Sequence 7892, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7892  
; LENGTH: 1217  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-7892

Query Match 30.7%; Score 59; DB 4; Length 1217;  
Best Local Similarity 41.7%; Pred. No. 7.2;  
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 12 GPGEFKTLRVQNLGCVVSGLI 35  
Db 1010 GIGPFSCIFRIQNLGLFPIHGMW 1033

RESULT 2

US-09-270-767-33880  
; Sequence 33880, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 33880  
; LENGTH: 469  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster

```
;
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-33880

Query Match      27.3%; Score 52.5; DB 4; Length 469;
Best Local Similarity 34.4%; Pred. No. 21;
Matches 11; Conservative 6; Mismatches 6; Indels 9; Gaps 2;

Qy 1 YEVHPYGTLPVG-----PQPEFKTTLRVQNLGC 28
Db 174 YLHPFSLPIGSRKPGPK-----RCREFXC 200

RESULT 3
US-09-270-767-49097
; Sequence 49097, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49097
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-49097

Query Match      27.3%; Score 52.5; DB 4; Length 469;
Best Local Similarity 34.4%; Pred. No. 21;
Matches 11; Conservative 6; Mismatches 6; Indels 9; Gaps 2;

Qy 1 YEVHPYGTLPVG-----PQPEFKTTLRVQNLGC 28
Db 174 YLHPFSLPIGSRKPGPK-----RCREFXC 200

RESULT 4
US-08-202-841A-2
; Sequence 2, Application US/08202841A
; Patent No. 6218596
; GENERAL INFORMATION:
; APPLICANT: Hughes, Stephen H.
; APPLICANT: Suttrave, Pramod
; APPLICANT: Pursel, Vernon
; TITLE OF INVENTION: Enhancement of Musculature in Animals
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,841A
; FILING DATE: 23-FEB-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/620,415
; FILING DATE: 03-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,449
```

```
;
; FILING DATE: 02-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/373,864
; FILING DATE: 30-JUN-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Alicea, Hector A.
; REGISTRATION NUMBER: 40,891
; REFERENCE/DOCKET NUMBER: 015280-170300US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 750 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 373
; OTHER INFORMATION: /product= "OTHER"
; OTHER INFORMATION: /note= "Xaa = Trp in c-ski;
; OTHER INFORMATION: Xaa = Arg in v-ski"
US-08-202-841A-2

Query Match      27.1%; Score 52; DB 3; Length 750;
Best Local Similarity 35.3%; Pred. No. 44;
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

Qy 10 PVGPGPEF-----KTTLRVQNLGCYVVG 33
Db 67 PVMPGPFMPSDRSTERCETILEGETISCFVVG 100

RESULT 5
US-08-356-354-2
; Sequence 2, Application US/08356354
; Patent No. 5767365
; GENERAL INFORMATION:
; APPLICANT: SONNEWALD, Uwe
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
; PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,354
; FILING DATE: 20-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/EP93/01605
; FILING DATE: 22-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P42 20 758.4
; FILING DATE: 24-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meilman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
```



;; INFORMATION FOR SEQ ID NO: 2;  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 846 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-356-354-2

Query Match 26.6%; Score 51; DB 1; Length 846;  
Best Local Similarity 35.5%; Pred. No. 71;  
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;  
Qy 1 YEHPYGTLPVGPGEFKTTLRVQNLGCVV 31  
Db 704 FKVCKPGTVP--PSKELKVMRIQALRCHAV 732

## RESULT 6

US-08-778-656-2  
; Sequence 2, Application US/08778656  
; Patent No. 5976869

;; GENERAL INFORMATION:  
;; APPLICANT: SONNEWALD, Uwe  
;; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE  
;; PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION  
;; NUMBER OF SEQUENCES: 6  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
;; STREET: 1180 Avenue of the Americas  
;; CITY: New York  
;; STATE: NY  
;; COUNTRY: US  
;; ZIP: 10036-8403

;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/778,656  
;; FILING DATE:  
;; CLASSIFICATION:

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/356,354  
;; FILING DATE: 20-DEC-1994  
;; APPLICATION NUMBER: US PCT/EP93/01605  
;; FILING DATE: 22-JUN-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: DE P42 20 758.4  
;; FILING DATE: 24-JUN-1992

;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Meilman, Edward A.  
;; REGISTRATION NUMBER: 24,735  
;; REFERENCE/DOCKET NUMBER: P/951-105  
;; TELEPHONE: (212) 382-0700  
;; TELEFAX: (212) 382-0888  
;; TELEX: 236925

;; INFORMATION FOR SEQ ID NO: 2;  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 846 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-778-656-2

Query Match 26.6%; Score 51; DB 2; Length 846;  
Best Local Similarity 35.5%; Pred. No. 71;  
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;  
Qy 1 YEHPYGTLPVGPGEFKTTLRVQNLGCVV 31  
Db 704 FKVCKPGTVP--PSKELKVMRIQALRCHAV 732

## RESULT 7

US-09-376-045-2  
; Sequence 2, Application US/09376045  
; Patent No. 6723898

;; GENERAL INFORMATION:  
;; APPLICANT: Aventis CropScience GmbH  
;; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF  
;; PLANTS WITH CHANGED SUCROSE CONCENTRATION  
;; FILE REFERENCE: 514413-3772  
;; CURRENT APPLICATION NUMBER: US/09/376,045  
;; CURRENT FILING DATE: 1999-08-17  
;; NUMBER OF SEQ ID NOS: 6  
;; SOFTWARE: Patent In Ver. 2.1  
;; SEQ ID NO 2

;; LENGTH: 846

;; TYPE: PRT

;; ORGANISM: Solanum tuberosum

US-09-376-045-2

Query Match 26.6%; Score 51; DB 4; Length 846;  
Best Local Similarity 35.5%; Pred. No. 71;  
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEHPYGTLPVGPGEFKTTLRVQNLGCVV 31  
Db 704 FKVCKPGTVP--PSKELKVMRIQALRCHAV 732

## RESULT 8

US-08-145-006C-5  
; Sequence 5, Application US/08145006C  
; Patent No. 5656452

;; GENERAL INFORMATION:

;; APPLICANT: Rao, Anjana

;; APPLICANT: Hogan, Patrick Gerald

;; APPLICANT: McCaffrey, Patricia

;; APPLICANT: Jain, Jugnu

;; TITLE OF INVENTION: NF-ATp, A T LYMPHOCYTE

;; TITLE OF INVENTION: DNA-BINDING PROTEIN

;; NUMBER OF SEQUENCES: 21

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Fish & Richardson P.C.

;; STREET: 225 Franklin Street

;; CITY: Boston

;; STATE: Massachusetts

;; COUNTRY: U.S.A.

;; ZIP: 02110-2804

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

;; COMPUTER: IBM PS/2 Model 50Z or 55SX

;; OPERATING SYSTEM: MS-DOS (Version 5.0)

;; SOFTWARE: WordPerfect (Version 5.1)

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/145,006C

;; FILING DATE: October 29, 1993

;; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 08/017,052

;; FILING DATE: February 11, 1993

;; APPLICATION NUMBER: 08/006,067

;; FILING DATE: January 15, 1993

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Fraser, Janis K.

;; REGISTRATION NUMBER: 34,819

;; REFERENCE/DOCKET NUMBER: 04590/007001

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (617) 542-5070

;; TELEFAX: (617) 542-8906

;; TELEX: 200154

;; INFORMATION FOR SEQ ID NO: 5:

;; SEQUENCE CHARACTERISTICS:

```
;
; LENGTH: 890
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-145-006C-5

Query Match 26.6%; Score 51; DB 1; Length 890;
Best Local Similarity 75.0%; Pred. No. 75;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 HPGTLPVGP 15
Db 805 HPLGTCVLP 816

RESULT 9
US-09-546-013-18
; Sequence 18, Application US/09546013
; Patent No. 6610504
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong-Shen
; TITLE OF INVENTION: METHODS FOR ASSAYING S-ADENOSYLMETHIONINE-DEPENDENT Methyltransfe
; FILE REFERENCE: 1937-1652
; CURRENT APPLICATION NUMBER: US/09/546,013
; CURRENT FILING DATE: 2000-04-10
; EARLIER APPLICATION NUMBER: 09/347,878
; EARLIER FILING DATE: 1999-07-06
; EARLIER APPLICATION NUMBER: 09/457,205
; EARLIER FILING DATE: 1999-12-06
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 18
; LENGTH: 890
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: T lymphocyte DNA binding protein: NF-At. sub.p
; PUBLICATION INFORMATION:
; PATENT DOCUMENT NUMBER: 5,656,452
; PATENT FILING DATE: 1993-10-29
; PUBLICATION DATE: 1997-08-12
US-09-546-013-18

Query Match 26.6%; Score 51; DB 4; Length 890;
Best Local Similarity 75.0%; Pred. No. 75;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 HPGTLPVGP 15
Db 805 HPLGTCVLP 816

RESULT 10
PCT-US94-00545-5
; Sequence 5, Application PC/TUS9400545
; GENERAL INFORMATION:
; APPLICANT: Rao, Anjana
; APPLICANT: Hogan, Patrick Gerald
; APPLICANT: McCallrey, Patricia
; APPLICANT: Jain, Jugnu
; TITLE OF INVENTION: NF-ATp, A T LYMPHOCYTE
; TITLE OF INVENTION: DNA-BINDING PROTEIN
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 555X

;
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00545
; FILING DATE: 18-JAN-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/145,006
; FILING DATE: October 29, 1993
; APPLICATION NUMBER: 08/017,052
; FILING DATE: February 11, 1993
; APPLICATION NUMBER: 08/006,067
; FILING DATE: January 15, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 04590/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 890
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; PCT-US94-00545-5

Query Match 26.6%; Score 51; DB 5; Length 890;
Best Local Similarity 75.0%; Pred. No. 75;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 HPGTLPVGP 15
Db 805 HPLGTCVLP 816

RESULT 11
US-08-356-354-6
; Sequence 6, Application US/08356354
; Patent No. 5767365
; GENERAL INFORMATION:
; APPLICANT: SONNEWALD, Uwe
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
; TITLE OF INVENTION: PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,354
; FILING DATE: 20-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/EP93/01605
; FILING DATE: 22-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P42 20 758.4
; FILING DATE: 24-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meilman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-105
```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 382-0700  
TELEFAX: (212) 382-0888  
TELEX: 236925  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 908 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-356-354-6

Query Match 26.6%; Score 51; DB 1; Length 908;  
Best Local Similarity 35.5%; Pred. No. 77;  
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEYHPYGTLPVGPGEFKTTLRVQNLGCVV 31  
Db 726 FKVCKPGTVP--PSKELKVMRIQALRCHAV 754

## RESULT 12

US-08-778-656-6  
Sequence 6, Application US/08778656  
Patent No. 5976869

## GENERAL INFORMATION:

APPLICANT: SONNEWALD, Uwe  
TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE  
PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
STREET: 1180 Avenue of the Americas  
CITY: New York  
STATE: NY

COUNTRY: US

ZIP: 10036-8403

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA: US/08/778,656  
APPLICATION NUMBER:  
FILING DATE:

## CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/356,354

FILING DATE: 20-DEC-1994

APPLICATION NUMBER: US PCT/EP93/01605

FILING DATE: 22-JUN-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE P42 20 758.4

FILING DATE: 24-JUN-1992

ATTORNEY/AGENT INFORMATION:

NAME: Meilman, Edward A.

REGISTRATION NUMBER: 24,735

REFERENCE/DOCKET NUMBER: P/951-105

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 382-0700

TELEFAX: (212) 382-0888

TELEX: 236925

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 908 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-778-656-6

Query Match 26.6%; Score 51; DB 2; Length 908;  
Best Local Similarity 35.5%; Pred. No. 77;  
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEYHPYGTLPVGPGEFKTTLRVQNLGCVV 31  
Db 726 FKVCKPGTVP--PSKELKVMRIQALRCHAV 754

## RESULT 13

US-09-376-045-6

Sequence 6, Application US/09376045

Patent No. 6723898

GENERAL INFORMATION:

APPLICANT: Aventis CropScience GmbH

TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF  
PLANTS WITH CHANGED SUCROSE CONCENTRATION

FILE REFERENCE: 514413-3772

CURRENT APPLICATION NUMBER: US/09/376,045

CURRENT FILING DATE: 1999-08-17

NUMBER OF SEQ ID NOS: 6

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 6

LENGTH: 908

TYPE: PRT

ORGANISM: Solanum tuberosum

US-09-376-045-6

Query Match 26.6%; Score 51; DB 4; Length 908;

Best Local Similarity 35.5%; Pred. No. 77;

Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEYHPYGTLPVGPGEFKTTLRVQNLGCVV 31  
Db 726 FKVCKPGTVP--PSKELKVMRIQALRCHAV 754

## RESULT 14

US-09-394-272-6

Sequence 6, Application US/09394272

Patent No. 6475588

GENERAL INFORMATION:

APPLICANT: Haigler, Candace H.

APPLICANT: Holaday, A. Scott

TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED  
EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE

FILE REFERENCE: 201304/1000

CURRENT APPLICATION NUMBER: US/09/394,272

CURRENT FILING DATE: 1999-09-10

NUMBER OF SEQ ID NOS: 14

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 6

LENGTH: 1053

TYPE: PRT

ORGANISM: Solanum tuberosum

US-09-394-272-6

Query Match 26.6%; Score 51; DB 4; Length 1053;

Best Local Similarity 35.5%; Pred. No. 91;

Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEYHPYGTLPVGPGEFKTTLRVQNLGCVV 31  
Db 911 FKVCKPGTVP--PSKELKVMRIQALRCHAV 939

## RESULT 15

US-08-356-354-4

Sequence 4, Application US/08356354

Patent No. 5767365

GENERAL INFORMATION:

APPLICANT: SONNEWALD, Uwe

TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE

PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
STREET: 1180 Avenue of the Americas  
CITY: New York  
STATE: NY  
COUNTRY: US  
ZIP: 10036-8403  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,354  
FILING DATE: 20-DEC-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/EP93/01605  
FILING DATE: 22-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P42 20 758.4  
FILING DATE: 24-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meilman, Edward A.  
REGISTRATION NUMBER: 24,735  
REFERENCE/DOCKET NUMBER: P/951-105  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 382-0700  
TELEFAX: (212) 382-0888  
TELEX: 236925  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1054 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-356-354-4

Query Match 26.6%; Score 51; DB 1; Length 1054;  
Best Local Similarity 35.5%; Pred. No. 91;  
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Cy 1 YEVHPYGTLPVGPDPFKTTLRVQNLGCVV 31  
: : | | | : : : : : : : : : :  
Db 912 FKUCKPGTVP--PSKELRKVMRIQALRCHAV 940

Search completed: April 6, 2005, 12:24:21  
Job time : 2.49374 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 39.4471 Seconds  
(without alignments)  
2761.097 Million cell updates/sec

Title: US-09-647-544-4  
Perfect score: 5932  
Sequence: 1 MEHPFVTHLPVFLVLTGLC.....GFFAHKKIPREKREKLEQ 1132

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:\*

1: Pirl:\*

2: Pirl2:\*

3: Pirl3:\*

4: Pirl4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1884	31.8	1180	2 A35854	integrin alpha-1 c
2	1891.5	31.7	1151	2 A45226	integrin alpha-1 c
3	1734.5	29.2	1170	2 I45914	integrin alpha-2 s
4	1733.5	29.2	1181	2 A33998	integrin alpha-2 c
5	1712.5	28.9	1178	2 S44142	VLA-2 protein homo
6	1115.5	18.6	1170	2 S03308	cell surface glyco
7	1105.5	18.6	1163	2 I56126	lymphocyte fuction
8	1091	18.4	1153	1 RWHU1B	cell surface glyco
9	1062.5	17.9	1163	1 RWHU1C	leukocyte surface
10	1025	17.3	1153	2 S00551	integrin alpha-B c
11	928	15.6	1179	2 A53213	integrin alpha-9 c
12	714.5	12.0	1035	2 I58409	alpaP integrin -
13	700	11.8	1054	2 JG7294	lymphocyte-Peyer's
14	666.5	11.2	1039	2 A41131	integrin alpha-4 c
15	665.5	11.2	1038	2 S06046	integrin alpha-1 -
16	655	11.0	272	2 A55348	integrin alpha-7 c
17	642	10.8	1137	2 JC5950	integrin alpha cha
18	637.5	10.7	1041	2 T31437	alpha-7 integrin -
19	609	10.3	1135	2 I61186	integrin alpha-3 c
20	593	10.0	1051	2 A40021	integrin alpha-v c
21	587	9.9	1044	2 T10050	integrin alpha cha
22	580	9.8	1106	2 S38783	integrin alpha-5 c
23	577.5	9.7	1048	2 A27421	integrin alpha-6 c
24	577.5	9.7	1091	2 A41543	integrin alpha-6 c
25	576	9.7	1072	2 B38457	integrin alpha-6 c
26	576	9.7	1073	2 A36429	integrin alpha-6 c
27	574	9.7	1053	2 I55534	VLA-3 alpha subuni
28	567.5	9.6	1045	2 S60571	integrin alpha v c
29	567	9.6	1044	2 S16516	integrin alpha-8 c

30	566.5	9.5	1049	2 A27079	fibronectin recept
31	564	9.5	1034	2 A36108	integrin alpha-V c
32	563.5	9.5	1053	2 S44250	integrin alpha-5 c
33	560	9.4	1051	2 A35761	cell surface glyco
34	534.5	9.0	1146	2 S40311	integrin - fruit f
35	528.5	8.9	1039	2 A34269	integrin alpha-2b
36	482.5	8.1	1037	2 A60163	glycoprotein IIB -
37	475	8.0	1394	2 A29637	position-specific
38	465.5	7.8	1226	2 S44824	F34F2.1 protein -
39	443	7.5	1139	2 S28277	hypothetical prote
40	407.5	6.9	1086	2 T18523	integrin alpha cha
41	405	6.8	1115	2 T09433	integrin alpha cha
42	405	6.8	1115	2 T09403	integrin alpha cha
43	394.5	6.7	191	2 T47230	VLA-2 protein - pi
44	362.5	6.1	764	2 I36916	glycoprotein IIB -
45	292	4.9	604	2 I36917	glycoprotein IIB -

ALIGNMENTS

RESULT 1

A35854

integrin alpha-1 chain precursor - rat

C/Species: Rattus norvegicus (Norway rat)

C/Date: 23-Oct-1990 #sequence\_revision 13-Sep-1991 #text\_change 09-Jul-2004

C/Accession: A35854; S11243

R/ignatius, M.J.; Large, T.H.; Houde, M.; Tawil, J.W.; Barton, A.; Esch, F.; Carbonetto,

J. Cell Biol. 111, 709-720, 1990

A/Title: Molecular cloning of the rat integrin alpha-1-subunit: a receptor for laminin a

A/Reference number: A35854; MUID:90338125; PMID:2380249

A/Accession: A35854

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-1180 <IGN>

A/Cross-references: UNIPROT:P18614; GB:X52140; NID:G56493; PIDN:CAA36384.1; PID:G56494

C/Keywords: cell adhesion; cytoskeleton; transmembrane protein

F:170-345/Domain: von Willebrand factor type A repeat homology <WAA>

Query Match 31.8%; Score 1884; DB 2; Length 1180;

Best Local Similarity 36.2%; Pred. No. 1.2e-131;

Matches 432; Conservative 219; Mismatches 432; Indels 112; Gaps 25;

Qy	13	LVFITGLCSFNLDEHPRLPPGPEAEFGYSVLQVGGGQRMVLGAPMGPSGDRGD	72
Db	19	LTVTILGFCVFNVDVKNSMSPSGPEVDMFGYVQQYENEEGKWLIGSLVGPQKARTGD	78
Qy	73	VYRCPCVGAHNAQCAKGLGDYQLGNSHPAV-----NMHLGMSILLETDDGGFMACAPL	127
Db	79	VYKCPVGRERAMPCKLDLP---VNTSIPNVTIENMTFG-STLVNPNNGGLACGPL	133
Qy	128	WSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVIVLDGNSNIYPWSEVQTF	187
Db	134	YAYRCGLHLYTTGICSDVSPTFQVNSFAP-VQECSTQLDIVIVLDGNSNIYPWSEVIAF	192
Qy	188	LRRVLGKFLIDPEIQVLQVYGSPVHWSLGFRTKEEVRAAKNLSRREGRETAKQ	247
Db	193	LNLLKRMIDGPKQTQVIGVQYGENVTHEFNLYKSYSTEELVAANKIGRQGLQTMAL	252
Qy	248	AIMVACTEGFSQSHGCRPEARLLVVVTDGESHGDEELPAALKACAGRVTRYGVAVLGH	307
Db	253	GIDTARKEAFTEARGARRGVKKVMVITVDGESHONYRLKQVIQDCEDENIQRFSAIILGH	312
Qy	308	YLRRQRDPSSPLRIRTIASDPDRFFNFVTDAAALTDIVDALGDRIFGLGSGHAENESS	367
Db	313	YNRGNLSTEFVEIKSIASEPTKEHFNVSDELATVIVKALGERIFALEATADQSAAS	372
Qy	368	FGLMSQIGFSTHRLKDGILLFGMVGVADWGSVLWEGGHLRFPFPRMALEDEFPALQNH	427
Db	373	FEMEMSGTGSFSAHYSQDWMLGAVGVDMNGTVMVQKANKQMPIHNTTTFOTE--PAKONE	430
Qy	428	--AAYLVGSYSMLLRGRRFLFLSGAPFRHRGKVIAPQLKKDGAVRVAQSLQGEIGSY	485

```
Db 431 PLASYLGTVNSATIPGD-VLYIAGQPRYNHTQGVYIKM-EDGNINILQTLGEGIGSY 488
Qy 486 FGSLCPLDTRDGTQTVLLVAAPMFLGPQNKETGRVVYLVGQSLTLTQGTLP----- 541
Db 489 FGSVLTITIDKQSYTDLLLVGAPMYGTKEBGKVIVYAV-NQTRFEYQMSLEPIRQT 547
Qy 542 -----EPPQDARFGAMGALPDLNQDGFADVAVGAPLEDGHHQGGALYLYHG 586
Db 548 C8SLKONSCYKKNKPEPCGARFOTAAVKDLNVDFNDVIGAPLEDDHAGAVIYHG 607
Qy 587 TQSGVRPHQRIAAAMPHALSIFGRSVGRDLDDGGDLVDVAVAGQAAILSSRPV 646
Db 608 SGTITREAYAQIRPSGGDKTLKFFGQSIHGMELNDGLTDTVTIGLGAALFWARDVA 667
Qy 647 HLTPSLEVTQAISVQDRCRRQEAVALCTAALCFQVTSRTPCRWHDQFVMTASLDE 706
Db 668 VVKVTFMFEKNVQKKNCRVBEKTVCTINATCFHVKLSKEDSIYEADLQIRVTLDS 727
Qy 707 WTAGARAAFQSGQRLSPRLRLSVGNVTCBQLHFHVLDTSDYLRLPVALTVTFALDNTTK 766
Db 728 LRQISRFSGTQERKIQR--NITVRESECIHRSFYMLDKHDFQDSVRVTLDF---NLTD 782
Qy 767 P--QPVNLBSPGTSIQKLPPSKDQDNECVTLVQLQNMDIRGSKAPFVVRGGRKV 824
Db 783 PENGVLDDALPNSVHEHIPPACDKGKERCISDLTLNVST---TEKSLLIIVKSQDKF 838
Qy 825 LVSTTLNRKENAYNTSLISIFSNLHLASITP-QRESPIKVECAAPSAHARLCSVGHV 883
Db 839 NVSLTVKNGDSAYNTTVQHSNLIIFSGIEEIQKDS-----CESNQITCRVGYPF 891
Qy 884 FQTGAKVTFLEFBSCLLSQVFGKLTASSDSLERNGLTQENTACTSAVIOYEPHLLF 943
Db 892 LRAGETVTKIIPQFNTHLSSENAIHLSATSDSEEPLESINDNEVNIPIPVKEVGLQF 951
Qy 944 SSBSTLHRYVHPVGTLP-----VG-----PGPFEKTL----- 972
Db 952 YSSASEHHSVAANETIEFINSTEDICNEINVEYTIKRGHFPMPBELQLSLSPNLTD 1011
Qy 973 -----RTNNAACIVQNTTEPPG-----PPVHPELQHTNLNGSNTOCQVRC 1015
Db 1012 GYPVLVPIGWSSSDNVNCRPSLSDPFGINSKGKQKTSKSEVLKRGITQDCSSTCGVATI 1071
Qy 1016 HLGOLAKG-TEVSVGLLRVLHNEEFRRRAKPSLTVVSTFELGTBEGSVLQTLTEASRWSES 1074
Db 1072 TCSLPSLDSQVNVL--LLWKPTFIRAHFSSNLTLRGLKXSNSS-LTSLSSNRKREL 1128
Qy 1075 LLEVQVT-RPILISLWILIGSVLGGLLALLVFLWKLWGPFAHKKIPBEEKRE 1128
Db 1129 AIQISKDGLPGRVPLWILLSAFAGLLMLLILALWKIGPF---KRPLKKMEK 1180

RESULT 2
A:5226
Integrin alpha-1 chain - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A45226
R:Briesewitz, R.; Epstein, M.R.; Marcantonio, E.E.
J. Biol. Chem. 268, 2989-2996, 1993
A>Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.
A:Reference number: A45226; MUID:93155124; PMID:8428973
A:Accession: A45226
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1151 <BRI>
A:Cross-references: UNIPROT:P56199
A:Experimental source: hepatoblastoma cell line HepG2
A>Note: sequence extracted from NCBI backbone (NCBIP:124326)
P:142-317/Domain: von Willebrand factor type A repeat homology <VWAL>

Query Match 31.7%; Score 1881.5; DB 2; Length 1151;
Best Local Similarity 35.9%; Pred. NO. 1.8e-131;
Matches 426; Conservative 220; Mismatches 423; Indels 119; Gaps 23;
```

```
Qy 23 FNLDERHPRLPFPPEAEFGYSVLQHVGGQRMVLVAPMDGPGSGDRRGDGVYRCPVGAH 82
Db 1 FNVDVKNMTFSFVDMFGYTVQQYENEBGKWLIGSPLVGQPKRTGDVYKCPVGRGE 60
Qy 83 NAPCAKGLHDYQLGNSHPAV-----NMHGLMSLLLETGDDGFMACAPLWSRACSSVF 137
Db 61 SLPCVKLDLP-----VNTSIPNVTEVKENMTFG-STLVTPNPGGFLACGPLYAYRCGHLHY 115
Qy 138 SSGICARVDASFOPOGSLAPTAQRCPTYMDVTVILGDSNSIYPWSEVQTLRLVLGKLF 197
Db 116 TTGICSDVSPFQVANSIAP-VQECSTQLDIVILGDSNSIYPWDSVTAFLNDLLKRMDI 174
Qy 198 DPQIOVGLVOYGESPVHWSLGDPRTKBEVRAAKNLRRREGRETAKAIWVACTEGF 257
Db 175 GPQQTQVGIQYQENVTHEFNLYKSYSTEELVAAKIVQRGROQTWALTGTDTARKEAF 234
Qy 258 SQSHGQRPEARLLVVVTDGESHGDBELPAALKACBAGRVTTRYGIAVLGHYLRQRDPSS 317
Db 235 TEARGARRGVKVMVITDGSNDHRLKKVIOQCEDENIQRFSAIILGSYNRGNLSTEK 294
Qy 318 FLREIRTIASDPDRPPFNTVDEAALTDIVDALGDRIFGLEGSHAENESSFGLMSQIGF 377
Db 295 FVEEIKSIASEPTEKHFNFVSDDELALVTIVKTGERIFALEATAADQSAASFEMMSQTGF 354
Qy 378 STHRLKDGILFGVMGAYDMGGSVLMLEGGHRLFPRLMALEDEFPALONHAAVLYGVSS 437
Db 355 SAHYSQDMVLMGAVGYDMNGTVVMQKASQIIPRNTTFNVESTKKNPELASVLYGVNS 414
Qy 438 MLARGRRRLFLSGAPRHRHGKVIAPQLKDGAVRVAQSLQGBQIGSYFGSSELCPLDTR 497
Db 415 ATASSGDVLIAGQPRVNTGVIIYRM-EDGNIKILQTLGSGQIGSYFGSILTTTIDK 473
Qy 498 DGTVDULLVAAPFLGPONKETGRVTVYLVGQSSLLTQGTLP----- 541
Db 474 DSNTDILLVAGPMYGTKEEGQKVYVAL-NQTRFEYQMSLEPIKQTCSSSRQHSCTT 532
Qy 542 ---EPPQDARFGAMGALPDLNQDGFADVAVGAPLEDGHHQGGALYLYHGTQGVVRHPAQ 598
Db 533 ENKNEPCGARFGTAIAVKDLNDGFNDIVIGAPLEDHGGAVIYHGSKTRKEVAQR 592
Qy 599 IAAASMPHALSYFGRSVDGRDLDDGDLVDVAVGAQAAIILSSRPVHLTPSLVTPQA 658
Db 593 IPSGGDKTKLFFGQSIHGMELNDGLTDTVTIGLGAALFWSRDVAVVVKVMTNPEPK 652
Qy 659 ISVQDRCRRQEAVALCTAALCFQVTSRTPCRWHDQFVMTASLDEWTAGARAAFDGS 718
Db 653 VNIQKNCHMEGKETVCINATVCFEVKLSKEDTIYEADLQYRVTLDSLRLQISRSPFSGT 712
Qy 719 GQRLSPRLRLSVGNVTCBQLHFHVLDTSDYLRLPVALTVTFALDNTTKP--GPVLNBSGP 776
Db 713 QERKVR--NITVRKSECTKHSFYMLDKHDFQDSVRITLDF---NLTDPEGVPVLDLSP 767
Qy 777 TSIQKLVPPSKDCGPNNECVTLVQLQNMDIRGSKAPFVVRGGRKVLVSTTLNRKEN 836
Db 768 NSVHEYIPPAKDCGKCEKISDLSLV-----ATTEKOLLIVRSQNDKFNVSILTVKNTKDS 823
Qy 837 AYNYSLSIIFSRNLHLASLTPQRESPIKVECAAPSAHARLCSVGHVFPQCAKVTFLLEP 896
Db 824 AYNTRTIVHYSPLNVFSGI-----EAIQKQSC--ESNNHNTCKVGYPLRREGVMVTFKILF 877
Qy 897 EFCSSLSLQVFGKLTASSDSLERNGLTQENTACTSAVIOYEPHLLSSSTLHRYBVHP 956
Db 878 QFNYSYLMENVTVILSATSDSEBPPETLSNVNVSIPVKYEVGLQYSSASEVHISIAA 937
Qy 957 YGTLP-----VG-----PGPEFKTLRLTNNASCIQVNLTEPPGPPVH 993
Db 938 NETVPEVINGSTEDIGNEINIFYLIRKSGSPMPBELKLSIS-----FPNMTSNGYPVLY 990
Qy 994 PEELOHTNRLN-----GSNTQCO-----VVRCHLQOLA 1021
Db 991 FTGLSSSENANCRPHIFEDPFSINSKGKMTTSTDLHKGRTILDNTCKFATITCNLTLS-S 1049
```

```
Qy 1022 KGTVEVCLLRLVHNEPRRAKFKSLTVVSTFELGTEGSLVQLTTEASRWSESLLEVQOT 1081
Db 1050 DISQNVSL--ILWKPTFKSYFSLNLTIRGELSENAS-LVLSNNQKELAIOISKD 1106
Qy 1082 -RPILISLWILGVLGGLLALLVFLCWLKGLGFAHKKIPEEEKREE 1128
Db 1107 GLPRVPLWILLSAFAGLALLMLLILALWLGIF--KRPLKKMEK 1151

RESULT 3
I45914
Integrin alpha 2 subunit - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 09-Jul-2004
C:Accession: I45914
J:Kamata, T.; Puzon, W.; Takada, Y.
J: Biol. Chem. 269, 9659-9663, 1994
A:Title: Identification of putative ligand binding sites within the I-domain of integrin
A:Reference number: A54402; MUID:94193647; PMID:7511592
A:Accession: I45914
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1170 <RAM>
A:Cross-references: UNIPROT:P53710; GB:L25886; NID:9439695; PIDN:AAB59255.1; PID:g439696
P:161-336/Domain: von Willebrand factor type A repeat homology <WMA>

Query Match 29.2%; Score 1734.5; DB 2; Length 1170;
Best Local Similarity 33.8%; Pred. No. 1.6e-120;
Matches 405; Conservative 229; Mismatches 456; Indels 109; Gaps 29;

Qy 9 LFLPLVLTGI---CSPNLDEHPRLPFGPEAFGYSVLQHVGGGQRMWLVGAPWDGP 65
Db 2 LQLVLVFSQGLNCCVAYNVGLPKAKIFSGSPSQFGYAVQVQFINPKGNWLLVGSWSGF 61
Qy 66 SGDRRGDVYRCPVCGAHNAPCAKGHIL-GDYOLGNSHHPAVNMHLCMSLELTDGQGFMAC 124
Db 62 PKRMGDVYKCPV-DLSTTTCENLQSTSMNSVNTKMTNWSLGLTLTRNVGTGGFLTC 120

Qy 125 APLMSRACGSSVFSGICARVDASFPQGSIAPTAQRCTYMDVYVLDGNSNTPMSEV 184
Db 121 GPLVAQCGSQYTTGVCSDVSPDFQLRTSPAPAVQTCPSFIDVVVVCDSNSIYPMDAV 180
Qy 185 QTFRLRLVKLFIPEQIOVLGVGESPVHWSLGDPTKEEVVRAAKNLSRREGRETK 244
Db 181 KNFLEKFOGDIQGTQOMGLIOYANNPRVVFNLNTPKSDKEMIKATQSTFOYGGDLTN 240
Qy 245 TAOAIMVACTGFGSOSHGRPEARLLVVVTDGSHDGEELPAALKACEAGRVTRYGIATV 304
Db 241 TFKAIOYARDATYSTAGRGFGATKVMVVVTDGSHDGSKUKAVIDQCCKNIILRFGIATV 300

Qy 305 LGHYLRQRDPSSFLREIRTIASDPDERFFNFVTDEAALTDIVDALGRIFLGEBSHAEN 364
Db 301 LGYLNRLALDTKNLIKEIKAIASIPTERHFNVSDEADLLEKAGTIGSQIFSIEGT-VQG 359
Qy 365 ESSFGLEMSQIGFSTRHKDQ--ILFGMVGAYDWGGSVLW-LEGHRLFPFRMALEDEFP 421
Db 360 GDNFQMEMSQVGSFAEYSPQNILMLGAVGAYDMSGTVOKTPHGLIFS-----KQAFE 414
Qy 422 PALQ--NHAAYLGVSVSSMLLRGRRFLSGAPFRHRGKVIAPOLKKGAVRVAQSLQOG 479
Db 415 QILOQRNHSYLGYSVAS-ISTGNSVHEVAGAPRANYTGQIVLVSNGENVNTVIQSQRG 473

Qy 480 EQIGSYFGSELCPDTRDGTDTDLLVAAPMFLGPQNKETGRVYVYLVGQOSSLTLTQOTL 539
Db 474 DQIGSYFGSVLCAVDVKNKDTITDLLVGCAPMYMNDLKKEGRVYVLTITKILANWHOPLE 533
Qy 540 QPEPPQDARFGAMALPDNLQDGPADVAVGAPLEDGHQGALYLYHGTQSGVRPHPAORI 599
Db 534 GPNGLNARFGSAALSDINMDGNDVIVGSPLENQNSGAVIYINHEGMIRLRYSOXI 593
Qy 600 AAASMPHA--LSYFGRSVDGRDLDDGLVDVAVGAQGAAILLSRPVHLTPSLVTPQ 657
Db 594 LGSDFAFSSHLQYFGRSLDGYDGLNGDSITDVSVGAGQGVVOLWSQSADVSDFTPK 653
```

```
Qy 658 AISVVQDCRRRGOEAVCLTAALCFQVTSRTPGRWDHQFYMRFTASIDEWTAG-----ARA 713
Db 654 KITILLNKAE-----IKLCLFSAKFR-PTNQNQVAIVYNTITDEDOFSRVSIRG 704
Qy 714 AFDGSGORLSPRRLRLSVGNVTCBQLHFVLDLTDYLRPVALVTYFALDNTTKPG--PVL 771
Db 705 LFKENNERCLOKTMIVSQAO-RCSEYIIHQPSDDIISPLMLCNWISLEN---PGTNPAL 760
Qy 772 NEGSPTSIQLVFPFKDCGPDNECVTDLVQLQNMNDRGSRKAPFVVRGRKRKVLVSTTLE 831
Db 761 EAYSETVKVFSIPPHKCGDGDGVCISDLVLNV-QQLPATQCOQPFIVSNQNKRLTFSVQLK 819
Qy 832 NRKENAVNTSLSIIFSRLHLASLTPORESPI-----KYEC-AAPSAHARLCSVGHVPFTG 887
Db 820 NKESAYNTEIVVDSENLFFASWS-----MPVDGTEVTCQIASQKSVTCNVGYPALKSK 875
Qy 888 AKVTFLEFFSCSLSSQVFGKLTASSDSLSRNGTLQENTAQTSAYIOYEPHLLFSSES 947
Db 876 QQVTFITNFENLQNLQNASISFRALSQSEEN--MADNSVNLKLSLLYDAEIHITRST 933
Qy 948 TLHRYEVHPYGTL-----PVGPGPEFKTTLRTNN-----ASCIVQ--NLTEPPGP-- 990
Db 934 NINFEVSLDGNVSSVHVSFEDIGPKFISIKVTGSGVPVSMASVIIHIPOYTKDKNPLM 993
Qy 991 -----PVHP-----BELQHNRLNGSNTQCVVRCHLQ 1019
Db 994 YLTGVHTDQAGDISCEAEINPLKIGQTSSSVSFKSENFRHIKELNCRTASCNSIMCWLRD 1053
Qy 1020 LAKGTEVSVGLLRLVHNEPRRAKFKSLTVVSTFELGTEGSLVQLTTEASRWSESLLEV 1079
Db 1054 LQVKGTEVFLNVSTRINWGTFAASTFTQVLTAAAEIDTNPQIYVIBE-----NTVTIPLT 1109
Qy 1080 QTRP---ILISLWILGSLVGLGLLALLVFLCWLKGLGFAHKKIPEEEKREEKL 1130
Db 1110 IMKPEKVEVPTGVIVGSVIAGILLALLVAILLWKLGLGFFKYEKMAKNPDETDETTTEL 1168

RESULT 4
A33998
Integrin alpha-2 chain precursor - human
N:Alternate names: CD49b; platelet glycoprotein GPIa; VLA-2/collagen receptor alpha-2 ch
C:Species: Homo sapiens (man)
C>Date: 30-Mar-1990 #sequence_revision 18-Sep-1992 #text_change 09-Jul-2004
C:Accession: A33998; B56793; A53117
R:Takada, Y.; Hemler, M.E.
J: Cell Biol. 109, 397-407, 1989
A:Title: The primary structure of the VLA-2/collagen receptor alpha(2)-subunit (platelet
A:Reference number: A33998; MUID:89308879; PMID:2545729
A:Accession: A33998
A:Molecule type: mRNA
A:Residues: 1-1181 <TAK>
A:Cross-references: UNIPROT:P17301; GB:X17033; NID:g33906; PIDN:CAA34894.1; PID:g33907
A:Note: The authors translated the codon GAT for residue 802 as Gln, GTC for residue 803
R:Caicmel, B.; Parentier, S.; Leung, L.L.; McGregor, J.L.
Biochem. J. 279, 419-425, 1991
A:Title: Separation of important new platelet glycoproteins (GPIa, GPIc, GPIIb, GPIIc and GPIIb
A:Reference number: A56793; MUID:92061944; PMID:1953640
A:Accession: B56793
A:Molecule type: protein
A:Residues: 30-43 <CAR>
R:Experimental source: platelet
R:Zutter, M.M.; Santoro, S.A.; Painter, A.S.; Teung, Y.L.; Gafford, A.
J. Biol. Chem. 269, 463-469, 1994
A:Title: The human alpha-2 integrin gene promoter. Identification of positive and negati
A:Reference number: A53117; MUID:94103255; PMID:8276836
A:Accession: A53117
A:Molecule type: DNA
A:Residues: 1-16, 'V', 18-21 <ZUT>
A:Cross-references: GB:L24121; NID:g400342; PIDN:AAA16619.2; PID:g4583535
A:Note: authors translated the codon GTA for residue 17 as Leu
C:Genetics:
A:Gene: GDB:ITGA2; CD49B
```

[illegible]

873 HARLCSVGHVPQTGAKVTFLLBFFPSCSLLSQVFGKLTASSDSLRNQTLOENTQAQTS 9332  
872 KSVACDVGYPALAKREQQVTFINFDNLQNLQNAQSLFQALSESQENKA--DNLVNLK 929  
933 AYIOYEPHLLFSSSESTLHRVEVHPYGTLP-----VQGPGRKTLTARTNNASCIVQ----- 982  
930 IPLYDAEIHILTRSTNINFYEISDGNVPISVHSFEDVDGPKFIISKVKTGTVGSPVPSMATV 989  
983 ----NUTEPPGP-----PVHP-----EELQHTNRLN 1004  
990 IIHIPQYTKENPLMWLTGVQTDKAGDISCNADINPLKIQTSSTSSVFKSENFRHTKELN 1049  
1005 GSNTQCOQVCHLQGLAKGTESVGLLRVLVHNEFFRRAKPKSLTVVSTFBLGTBEGSVLQ 1064  
1050 CRTASCSNVTCWLKDVHMKGEYFNVVTRTWNGTFASTPQTQVLTAAABINTYNPEIY- 1108  
1065 LTEASRWSELSLEVQTRPLISLWI-----LIGSVLGGLLALLLVPLCLWK 1111  
1109 -----VIDNTVTIPLMKPDEKAEVPTGVTIGSTIAGILLALLLVAILLWK 1155  
1112 LGFFFAHK-----KIPPE 1123  
1156 LGFFPKRYEKMTKNPDE 1172

RESULT 5  
S44142  
VLA-2 protein homolog - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004  
C:Accession: S44142  
R:Edelman, J.M.; Chan, B.M.; Uniyal, S.; Onodera, H.; Wang, D.Z.; Damjanovich,  
submitted to the EMBL Data Library, January 1994  
A:Description: The mouse VLA-2 homologue supports collagen and laminin adhesion  
A:Reference number: S44142  
A:Accession: S44142  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1178 <EDE>  
A:Cross-references: UNIPROT:Q62469; EMBL:Z29987; NID:g473098; PIDN:CAA82877.1;  
F:169-344/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match 28.9%; Score 1712.5; DB 2; Length 1178;  
Best Local Similarity 34.7%; Pred. No. 7.3e-119;  
Matches 412; Conservative 217; Mismatches 457; Indels 103; Gaps 32;

9 LFLPLVLFT--GL--CSPFNLDHHRPLRPPPPAEFGYSVLQHVGGQRWMLVGAPWDGP 65  
10 LLLQLMLVQGLINCLAYNVLGPKAIFSGPSSEQFGYSVQQLTNPQGNWLLVGSPPSGF 69  
66 SGDRRGDVYRCPVGGAHNAPCAKGHLL-GDYQLGNSSHPAVNMHLGMSLLETDDGGGPMAC 124  
70 PENRMGDVYKCPV-DLPATCEKLNQLNSASISNVTEIKTNMSLGLTLTRNPGTGGPLTC 128  
125 APLWSRACGSSVSSGSCARVDASFPQGS LAPTAQRCPYMDVVIVLDCSNSIYPMSEV 184  
129 GPLMAHQCNGQYATGICSDVSPDFQLTFSFPAVQACPSLADVWVVCDESNSIYPMSEV 188  
185 QTFURRLVKGKLFIDPEIQVGLVQSGSPVHVESGLGDFRTKEEVVRAAKNLSRREGRETK 244  
189 KNFLVKFTVGLDIGPKTQVALIQYANEPRILFNLDNFKEDMVQATSETRQHGGLDNLN 248  
245 TQAQIMVACTEGFSQSHGGRPEARLLVVTVDGESHGGBELPAAKACAGRVTRYGIIV 304  
249 TFRAIREFARDYAYSQTSGGRPFGATKVMVWVTDGESHGDKLKTVTIQCNDDIELRFGIIV 308  
305 LGHYLRQRDPSSPLRIIRTIASDPDRFPFNVTDEAALTIDVLDALGDRIFGLEGSHAEN 364  
309 LGYLNRLALDTKNILKEIKAIASPTBRYIFNVADAELALLEKAGTLGEQIFSLTGT-VQG 367  
365 ESSFGLMSIQFGFSTHRL--KDGILFGMVQAYDWGGSVLMLLEGHR--LFFPPRMALEDEF 420



Db 368 GDNFQMEQAQVGFADYAPQNDILMLGAVGAFDMSGTLV-QETSHKPVIP- - - - -KQAF 421  
Qy 421 PPALQ--NHAAYLYGVSSVSMMLRGRRLFLSGAPRRHRGKVIAPOLKKGAVVAQSLQ 478  
Db 422 DOVLQDRNHSFLGSVAAISTEDGVH-FVAGAPRANTGQIVLYSVNKGQNVTVIQSHR 480  
Qy 479 GEQIGSYFGSLCPDTRDGTDTDLVAAPMELGPQNKETGRVYVVLVGGQSLTLTQGT 538  
Db 481 GDQIGSYFGSLVCSVDVDKOTITDVLVGAFTYNDLKEGKVFYFTITKILNQHQFL 540  
Qy 539 LQPEPPQDARFGFANGALPDNLQDGFADVAVGAPLEDHQALYLYHGTSGVGRPHPAQR 598  
Db 541 EGPEGTGNARFSAIALSDINMGDFNDVIVGSPVENENSGAVIYNGHQTITKYSQK 600  
Qy 599 IAAA--SMPHALSFGSRVGRDLDDGLVDVAVGAGAAIILSSRPVHLTSPLEVTP 656  
Db 601 ILGNGAPRRHLQFGRSLDGYGLNGDSITDVSIGALQVQIQLWSQSIADVAEALFTP 660  
Qy 657 QAISVQDRCRRQGEAVCLTAALCFQVTSPTGRWDHOFYMRFTASIDEWTAGARAAD 716  
Db 661 DKITLLNKDAK-----ITLKCFRAEFPAQO--NNQVAILFNMTLDADGHSRVTSR 711  
Qy 717 GSGQRLSPRRLR--LSVGNV--TCQLHPHVLDTSDYLRPVALVTTFALDNTTKPG--PVL 771  
Db 712 GVFRENSERFLQKNMVYVQKCEHHISIQKPSDVNPLDLRVDISLEN---FGTSPAL 768  
Qy 772 NEGPTSQIKLVPSKCGPNECVTDLVLQVNDIRGSRKAPFVVRGRRKVLVSTTLE 831  
Db 769 EAYSETVKVFSIPFYKEGCSGICITSDILDV-QOLPAIQTSFIVSNQNKRLTFSVILK 827  
Qy 832 NRKENAVNTSLIIFSRNLHLASITPQRESPI---KVECAPSAHARL-CSVGHVPVQTG 887  
Db 828 NRGESAVNTVLAEFSENLFASFPS---MPVDGTEVTCVSGSQKSVCTDVGVPALKSE 883  
Qy 888 AKVTFLEFEFPCSSLLSQVEKGLTASDSLSERNGLTQENTAQTSAYTOVEPHLLFSSES 947  
Db 884 QOVTTFINFDNLQNLQNAINQAFSESQETNKA--DNSVSLTIPLYDAELHLTEST 941  
Qy 948 TLHRYEHPYGTLP-----VQFG-----PEFK----- 969  
Db 942 NINFYEISSDENAPSVIKSVEDIGPKFTFLSVKTAGSAPVSMALVTIHIPOYTKENPLL 1001  
Qy 970 --TTLRTNNA--SCI--VQNLTEP---PGPVHPELOHTNRLNGSTQCVVRCHLQ 1019  
Db 1002 YLTGIQTDQAGDISCTABINFLKLPHTAPSFSFKNENFRHTKELDCRTTSCNTLWLD 1061  
Qy 1020 LAKTEVSVGLRLVHNEFRFRKSLTVVSTFELGTEEGSVLQLEASRWSLSLEV 1079  
Db 1062 LHMAEYFINTVTRVWRTFAASTFQTVQLTAAAEIDTHNPQ-LFVIEENAVTIPLIMK 1120  
Qy 1080 QTRPILISLWILGSLVGLGLLLALLVFLMKLQFF--AHKKI---PEE 1123  
Db 1121 PTERAEVPTGVIIGSIAGIILLAMTAGLWKLGFGRKQYKMGQNPDE 1169

RESULT 6  
S03308  
cell surface glycoprotein CD11a precursor - human  
N;Alternate names: leukocyte adhesion glycoprotein LFA-1 alpha chain; leukocyte function  
C;Species: Homo sapiens (man)  
C;Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 09-Jul-2004  
C;Accession: S03308; A47458; A47565; A48759; S36044  
R;Larson, R.S.; Corbi, A.L.; Berman, L.; Springer, T.  
J. Cell Biol. 108, 703-712, 1989  
A;Title: Primary structure of the leukocyte function-associated molecule-1 alpha subunit  
A;Reference number: S03308; MUID:89139587; PMID:2537322  
A;Accession: S03308  
A;Molecule type: mRNA  
A;Residues: 1-1170 <LAR>  
A;Cross-references: UNIPROT:P20701; UNIPROT:Q9UBC8; EMBL:Y00796; NID:G31421; PIDN:CAA687  
A;Note: Part of this sequence was confirmed by protein sequencing  
R;Cornwell, R.D.; Gollan, K.A.; Hickstein, D.B.  
Proc. Natl. Acad. Sci. U.S.A. 90, 4221-4225, 1993

A;Title: Description of the leukocyte function-associated antigen 1 (LFA-1 or CD11a) pro  
A;Reference number: A47458; MUID:93248261; PMID:8097887  
A;Accession: A47458  
A;Molecule type: DNA  
A;Residues: 1-20 <COR>  
A;Note: sequence extracted from NCBI backbone (NCBIN:130862, NCBI:P130863)  
R;Shelley, C.S.; Farkhzad, O.C.; Arnaout, M.A.  
Proc. Natl. Acad. Sci. U.S.A. 90, 5364-5368, 1993  
A;Title: Identification of cell-specific and developmentally regulated nuclear factors t  
A;Reference number: A47565; MUID:93281759; PMID:8099450  
A;Accession: A47565  
A;Molecule type: DNA  
A;Residues: 1-20 <SHE>  
A;Cross-references: GB:M95609  
R;Nueda, A.; Lopez-Cabrera, M.; Vara, A.; Corbi, A.L.  
J. Biol. Chem. 268, 19305-19311, 1993  
A;Title: Characterization of the CD11a (alphaL, LFA-1alpha) integrin gene promoter.  
A;Reference number: A48759; MUID:93374910; PMID:8103515  
A;Accession: A48759  
A;Molecule type: DNA  
A;Residues: 1-20 <NUE>  
A;Cross-references: EMBL:Z22804; NID:G311405; PIDN:CAA80461.1; PID:G311406  
C;Genetics:  
A;Gene: ITGAL; CD11A  
A;Cross-references: GDB:119757; OMIM:153370  
A;Map position: 16p11.2-16p11.2  
C;Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo  
C;Keywords: cell adhesion; cytoskeleton; glycoprotein; heterodimer; surface antigen; tra  
F;1-25/Domain: signal sequence #status predicted <SIG>  
F;26-1170/Product: leukocyte adhesion glycoprotein LFA-1 alpha chain #status predicted <  
F;154-317/Domain: von Willebrand factor type A repeat homology <WAA>

Query Match 18.8%; Score 1115.5; DB 2; Length 1170;  
Best Local Similarity 30.0%; Pred. No. 2.3e-74;  
Matches 368; Conservative 191; Mismatches 454; Indels 213; Gaps 55;

Qy 11 LPLVFLTGL-----CSPFLNDEHHPLFPGPPEA--EFGYSVLQHVGGQRMVLVGPWD 63  
Db 9 MAMALLSGFFFPASSNLDVRGARSF-SPPRAGRHFGRVQLQ-VGNG---VIVGAPGE 63  
Qy 64 GPSDRDRGVYRCVPGVAHNAAPCAKGLGDYQGLGNSHPAVNMHLGMSLLETDCGGGMA 123  
Db 64 GNS---TGLYQCOSGTGCHLPVT-----LRGSNY--TSKYLGMTLATDPTDGSILA 110  
Qy 124 CAPLWSRACSSVSSFGICARVDASFO-PQGS LAPTAQRC-PTVMVIVVLDGNSIYP- 180  
Db 111 CDPGLSRTCDQNTYLSGLCYLFRNLQCPMLQGRPGQECIKGNVDLVFLPDGMSLQPD 170  
Qy 181 -WSEVQTFRLRLVGLKLFIDPEQIQVGLVQYGESPVHWSLGDF---RTKEBVVRAAKNLS 236  
Db 171 EFQKILDFMKDVMKKL--SNTSYQFAAVQFSTSYKTFDFSDYVYKXDPDALLKHVXHML 228  
Qy 237 RREGRETKTAQAIMVACTEGFSQSHGGRPEAARLLVVVTDGESHGDELPALKACEAGR 296  
Db 229 LL-----TNTGAINYVATEVEFREELGARPDATKVLIIITDGEATDSGNIDAKD----- 278  
Qy 297 VTRYGIAVLGHYLRQRDPSSFLEIRTIASDPDERFFNVNVTDEAALTDIVDALGDRIFG 356  
Db 279 IIRYIIGIKGHFQKESQET-----LHKFASKPASEFVKILDTPEKLDLFTLQKKIYV 333  
Qy 357 LEGSHAENESFGLMSQIGFSTHRLKDXGLFGVAYDNGCVSLWLEGGHRLFPFPMAL 416  
Db 334 IEGTSKQDLTSFNMELSSGISADLSRGHAVVGVAVGADWAGGFLDKADILQ----- 385  
Qy 417 EDEF-----PPALQNHAAVLYGVSSVSMMLRGRRLFLSGAPRRHRGKVIAPOLKKGAV-V 471  
Db 386 DDTFIGNEPITPEVRAGYLGTVTWLPSRQKTSLLASGAPRYQHWGRVLLFQEPQCGGHW 445  
Qy 472 RVAQSLQGEIGSYFGSELCPDTRDGTDLVAAPMELGPQNKETGRVYVY---LVG 528  
Db 446 SQVQTIHGTOIGSYFGGELGCVDDVDQGETELLIGAPLFYGEQ--RGRVFIYQRQLG 503  
Qy 529 QQSLLTLQGTLOPEPPQDARFGFANGALPDNLQDGFADVAVGAPLEDHQALYLYHGTQ 588

```
Db 504 FEEVSELQ--DPGYPL-GRFGEAITALTDINGDLVDVAVGAPLEE--QGVAVIFNGRH 558
Qy 589 SGVRPHPAQRTAAASMPHALSYFGRSDRLDLDGDDLDVAVCAQGAAILLSRPVHL 648
Db 559 GGLSPQPSQRLEGTVQLSGIQMFGRSHGVKXDLGGLADVAVGAESQMVLSSRPVDM 618
Qy 649 TPSLEVTPOAISVVQRDC---RRRQGEAVCLTAAALCFQVTSRTPPGRMDHOFYMR-----699
Db 619 VTLMSPPABEIPVHEVECSYSTSNKMEGVNIT--ICFQIKSLYP-----QFGRLVANL 671
Qy 700 -FTASLDWETAGARAAFDGSGQRSLRRLRLSVGNVTCBQLHFHV-LDTSYLRPVALTV 757
Db 672 TYTLQLDGHRTTRRGLFPGRHSL--RRNIAVTTSMSTDFSPHFPVCVDLSPINVS 729
Qy 758 TPAL---DNITKPGVNLNBSPTSIQKL-----VPSKDCGPDNECVTDLVQVNV-MD 806
Db 730 NFSLWBEGETPRDQRAQKQIPILRPSLSHSETWEIPFKNCGEDKKCEANLRVSPGAR 789
Qy 807 IRGSRKAPFVVRGRRKVLVSTLTLENKKNAYNTSLSIIFSRNLHLASLTQR-ESPIKV 865
Db 790 SRALRLTAF-----SLSVLSLSNLEEDAYWQLDLHFPGLSPFKVEMLKPHSQIPV 843
Qy 866 ECAAPSAHARL-----CSVGHVPVQFGAKVTFLLPEFSCSSLLSQVFGKLTASS--DS 917
Db 844 SCELPBESRLLSRALSCNVSSPIFKAGHSA-----LQMMFNTLVNSSWGDS 891
Qy 918 LERNGT-----LQENTAQTSAYIQVEPHLLF--SSESTL-----H 950
Db 892 VELHANVTNNEDSLLDNSATTIIPILYPINILIOQEDSTLVYVFTPKPGKIHQVKH 951
Qy 951 RYEV-----HPYGTL-----PVGPGPEKTLRTNNASCIVQNLTEPPGPPVHP 994
Db 952 MYQVRIQPSIHQHNIPITLAVVGVQPPSEGP-----ITHQWS--VQ--MEPP-VCHY 1000
Qy 995 BELQHTNRLNGSNTQC---QVVRCHLQGLAKGTVEVSGLRLV---HNEPFRRAKPKSL 1047
Db 1001 EDLE---RLPDAEPCLPFGALFRCVVPVFRQILVQVIGTLELVEIEASSMFLSCSLSI 1057
Qy 1048 TVVST--PELGTGSEGLVQLTEARWSESLLEVQVTRPILISILWILGSLGGLLALL 1105
Db 1058 SFNSKSHFLYGSNASLAQV-----MKVDVVEYKQMLY-LYVLGS--IGGLLLLLLI 1107
Qy 1106 VFCLWLKGFPAHKKIPBEKEEKE 1131
Db 1108 FIVLYKVGFP-----KGNLKEKE 1126

RESULT 7
I56126
Lymphocyte fuction-associated molecule-1-alpha - mouse
C/Species: Mus musculus (house mouse)
C/Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 09-Jul-2004
C/Accession: I56126
R/Kaufmann, Y.; Teeng, E.; Springer, T.A.
J. Immunol. 147, 369-374, 1991
A/Title: Cloning of the murine lymphocyte function-associated molecule-1 alpha-subunit
A/Reference number: I56126; MUID:91268576; PMID:2051027
A/Accession: I56126
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1163 <RES>
A/Cross-references: UNIPROT:P24063; GB:M60778; NID:g198785; PIDN:AAA39426.1; PID:g198786
C/Genetics:
A/Gene: LPA-1
C/Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homology
F;151-315/Domain: von Willebrand factor type A repeat homology <VMA1>

Query Match 18.6%; Score 1105.5; DB 2; Length 1163;
Best Local Similarity 28.8%; Pred No.1.3e-74;
Matches 359; Conservative 180; Mismatches 449; Indels 259; Gaps 49;

Qy 12 PLVFLTGL-----CSPFNLDHHPRLPFGPPEAFGYSLQHVGGQRMVLVGAPWDGSPS 66
```

```
Db 8 PRLLLLGLQLFAKAWSYNLDTRPTQSLAQAGRHFGVQLQIEDG---VVVAGPGE---60
Qy 67 GDRRGDVYRCPVCGAHNAPCAKGLHDYQIYGNSSHPAVNMHLGMSLLETDDGGPWCAP 126
Db 61 GDMWTGGLYHCRTSSEFCQPVVS-----LHGSNH--TSKILGMLTATDAKGSLLACDP 110
Qy 127 LWSRACSSSVFSGICARVDASFPQGS LAPTAQRCPY-----MDVVIVLDGNSI-178
Db 111 GLSRTCDQNTYLSGLC-----YLPQSLGPMQLNRPAYQECMKGVLDVLPFDGSGSLD 165
Qy 179 -YPMSEVQVTLRRLVLGKLFIDPQIQVGLVQYGESPVHWSLGDFTKKEVVRRAAKNLSR 237
Db 166 RKDFEKILEPMKDMVKL--SNTSYQFAAVQFSDCTCRTEFTLDY-----VKQNKNDV 217
Qy 238 REGRE-----TKTAQAINVACTEGESQSHGGPEARLLVVTVDGESHGDEBELPAALKA 291
Db 218 LLGSVQPMFLTTTFRAINVVAHVFKBESGARPDATKVLVIITDGEASDKGNISA---273
Qy 292 CEAGRVTTRYGIAVLGHYLRQRDPSSGFLREIRTIASDPDERFFNFVTDAAALTDIVDALG 351
Db 274 --AHDITRYIIGIKHFSVKQ-----KTLHIFASEPVEEFVKILDTPEKLDLFDLQ 326
Qy 352 DRIFLEGSHAENESSFGLEMSQIGFSTHRLKDGILFGVGAAYDMGGSVLMLEGHRLFP 411
Db 327 RRIYATBGTNRQDLTSPNMLSSGISADLSKGHAVVGAAGKDWAGGFLDLR-----379
Qy 412 PRMALED-----EPFPALQNHAAVLGYSVSSMLRGRRLLFLSGAPFRHKGVIAP 463
Db 380 -----EDLOGATVGOEPLTSDVRGGYLVTVAMTSSRSRPLLAAGAPRYQHVGQVLLF 434
Qy 464 QL-KKQCAVRAVQASLOQEQIGSYFGSELCLPDTDRDGTDTDLVLAAPMFLPQMKETGRV 522
Db 435 QAPAGGRWNOTKIEGTQIGSYFGELCSVDLDQGEAELLIGAPLFGEO--RGRV 492
Qy 523 YVYLVGQSSLLTLOGLTQPEPPQD--ARFGPAMGALPDLDQDGFADVAVAPLEBHGHAL 581
Db 493 FTY-QRRQSLFENVSELQDGPVPLGRFGAAITALTDLINGDRDLTDVAVGAPLEE--QGV 549
Qy 582 YLHGTOGVRPHPAQRIAAASMPHALSYFGRSDRLDLDGDDLDVAVCAQGAAILLS 641
Db 550 YIFNGKPGGLSPQPSQRIQGAQVFPGRWFRSHGVKIDLGDRDLADVAVGAEGRVVLS 609
Qy 642 SRPIVHLTPSLEVTPOAISVVQRDC---RRRQGEAVCLTAAALCFQVTSRTPPGRMDHOFY 697
Db 610 SRPVVDVTELSFSPBEIPVHEVECSYSAREEQKGVKLA--CFRIKPLTP-----QFQ 662
Qy 698 MR-----FTASLDWETAGARAAFDGSGQRSLRRLRLSVGNV-----TCBQLHFHV-L 744
Db 663 GRLLANLSYTLQDGHMRSRGLFPDGSHEL-----GNTSITPKDSCDPHFHFP 714
Qy 745 DTSYLRPVALTVTFALDNTTKPGVNLNBSPT-----SIQKLVPSKD 788
Db 715 CIQDLSPINVLNFSL-----LEEEGTPRDQKGRAMQPILRPSIHVTKEIPFKN 766
Qy 789 CGPDNECVTDLVQVNVMDIRGSRKAPFVVRGGRKVLVSTLTLENKKNAYNTSLSIIFSR 848
Db 767 CGEDKKEANLTL-----SPASGGLRLMSS--ASLAVETLNSGDEAVVRLDLDFPR 820
Qy 849 NLHLASLTP-QRESPIKVECAAPSAHARL-----CSVGHVPVQFGAKVTFLLPEFSCS 901
Db 821 GLSPRKVEMLQPHSRMPSVCEELTEGSSLLTKLKNVSSPIFKAGQEV-----870
Qy 902 SLLSQVFGKLTASS--DSLERNGLT-----QENTAQTSAYIQVEPHLLFSSE--S 947
Db 871 --LQVMFNTLVNSSWEDFVSLNGVHCENENSSIQEDNSAATHIPVLPVNLTKQENS 928
Qy 948 TLHRYEVHPYGTLPVGPSPFKTTLRTNNASCIVQNLTEPPGPPVHPPEELQHTNRLNGSN 1007
Db 929 TLY-----ISFPFKGP-----KTQQVHVYQVRIQ---PSAYDHNPTLEALVG--969
Qy 1008 TQCOVVRCHLQGLAKGT-EVSVGLLLRLVHNEFFRRAKFKS-----1046
```

Db 970 ---VPRHSEDLITYTWSVQTDPLVLTCHSDKRPSSSEARQPCLVGVQFRCPIVERWEI 1025  
Qy 1047 -LTVVSTFELGTE--EGSVQLQTEA-----SRNSES-----LLEVVQTRPI 1084  
Db 1026 LIQVTGVTSKELKASSTLSLCSLSVSFNSSKHFLYSGKASEAQVLKVDLIHKEM 1085  
Qy 1085 LISWILIGSVLGGLLALLVFLCWLKLGFFPAHKIPDEEKREKLE 1131  
Db 1086 L-HYVVLG--IGGLVLLFLFLALYKGVFF-----KRLKKEWE 1122

RESULT 8  
RWHLB  
cell surface glycoprotein CD11b precursor [validated] - human  
N;Alternate names: complement receptor type 3 alpha chain; leukocyte adhesion protein Ma  
eukocyte integrin alpha chain; neutrophil adherence receptor alphaM chain  
C;Species: Homo sapiens (man)  
C;Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text change 09-Jul-2004  
C;Accession: A31108; A28915; A41600; A30892; A32218; A46526; A26091; I52567  
R;Corbi, A.L.; Kishimoto, T.K.; Miller, L.J.; Springer, T.A.  
J. Biol. Chem. 263, 12403-12411, 1988  
A;Title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD  
B.  
A;Reference number: A31108; MUID:88315033; PMID:2457584  
A;Accession: A31108  
A;Molecule type: mRNA  
A;Residues: 1-1153 <COR>  
A;Cross-references: UNIPROT:P11215; GB:J03925; NID:G187284; PIDN:AAA59544.1; PID:G307148  
A;Note: part of this sequence was confirmed by protein sequencing  
R;Arnaout, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.  
J. Cell Biol. 106, 2153-2158, 1988  
A;Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor M  
A;Reference number: A28915; MUID:88257215; PMID:2454931  
A;Accession: A28915  
A;Molecule type: mRNA  
A;Residues: 1-499, 501-965, 'P', 967-1153 <ARN>  
A;Cross-references: GB:M18044; GB:J03270; GB:M19664; GB:X07421; NID:G186935; PIDN:AAA594  
A;Note: the authors translated the codon TAC for residue 1129 as Thr  
R;Shelley, C.S.; Arnaout, M.A.  
Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991  
A;Title: The promoter of the CD11b gene directs myeloid-specific and developmentally reg  
A;Reference number: A41600; MUID:92073318; PMID:1683702  
A;Accession: A41600  
A;Molecule type: DNA  
A;Residues: 1-9 <SHE>  
A;Cross-references: GB:M76724; NID:G180018; PIDN:AAA58410.1; PID:G553215  
R;Arnaout, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Harris, P.; Tenen, D.G.  
Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988  
A;Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesi  
A;Reference number: A94193; MUID:88190151; PMID:2833753  
A;Accession: A30892  
A;Molecule type: mRNA  
A;Residues: 917-1042 <AR2>  
A;Cross-references: GB:M18044  
R;Hicksrein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.M.; Back, A.L.; Roth, G.J.  
Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989  
A;Title: cDNA sequence for the alphaM subunit of the human neutrophil adherence receptor  
A;Reference number: A32218; MUID:89098893; PMID:2563162  
A;Accession: A32218  
A;Molecule type: mRNA  
A;Residues: 9-1153 <HIC>  
A;Cross-references: GB:J04145; NID:G189068; PIDN:AAA59903.1; PID:G386975  
A;Note: part of this sequence was confirmed by protein sequencing  
R;Fleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.F.; Tenen, D.G.  
J. Immunol. 150, 480-490, 1993  
A;Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-in  
n during evolution.  
A;Reference number: A46526; MUID:93123748; PMID:8419480  
A;Accession: A46526  
A;Status: not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-499,501-1153 <FLE>

A;Cross-references: GB:S52227; NID:G263047; PIDN:AAB24821.1; PID:G263049  
A;Note: the last three bases of intron 13, CAG, are included in some but not all mature  
A;Note: sequence extracted from NCBI backbone (NCBIP:121963)  
R;Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnaout, M.A.  
Biochim. Biophys. Acta 874, 368-371, 1986  
A;Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across sp  
A;Reference number: A90664; MUID:87076671; PMID:3539202  
A;Accession: A26091  
A;Molecule type: protein  
A;Residues: 17-31 <PIE>  
A;Experimental source: granulocytes  
R;Pahl, H.L.; Rosmarin, A.G.; Tenen, D.G.  
Blood 79, 865-870, 1992  
A;Title: Characterization of the myeloid-specific CD11b promoter.  
A;Reference number: I52567; MUID:92144986; PMID:1346576  
A;Accession: I52567  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-9 <RES>  
A;Cross-references: GB:M84477; NID:G180184; PIDN:AAA51960.1; PID:G553219  
C;Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1  
C;Genetics:  
A;Gene: GDB:ITGAM; CR3A  
A;Cross-references: GDB:120599; OMIM:120980  
A;Map position: 16p11.2-16p11.2  
A;Note: promoter contains a GATA motif and two Spl consensus binding sites  
C;Superfamily: cell surface glycoprotein CD1b; von Willebrand factor type A repeat homo  
C;Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer; mag  
F;1-16/Domain: signal sequence #status predicted <SIG>  
F;17-1153/Product: cell surface glycoprotein CD1b #status experimental <MAT>  
F;17-1108/Domain: extracellular #status predicted <EXT>  
F;148-318/Domain: von Willebrand factor type A repeat homology <VWA>  
F;465-473/Region: calcium/magnesium binding #status predicted  
F;530-538/Region: calcium/magnesium binding #status predicted  
F;593-601/Region: calcium/magnesium binding #status predicted  
F;1109-1134/Domain: transmembrane #status predicted <TM>  
F;1135-1153/Domain: intracellular #status predicted <INT>  
F;86,240,331,469,693,697,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Binding  
Query Match 18.4%; Score 1091; DB 1; Length 1153;  
Best Local Similarity 29.6%; Pred. No. 1.5e-72;  
Matches 359; Conservative 202; Mismatches 482; Indels 170; Gaps 47;  
Qy 11 LPLVFLTG--LCSFFNLDEHHPLFPGPPEAEFGYSLVGHVGGQRMWLYCAPWDGSGD 68  
Db 3 LRVLLLTALTLCGHFNLDTENAMTFQENARG-FQGSVVLQGSR---VVGAPQEIWAAN 58  
Qy 69 RRGDVYRCPVGAHNAFCAXGHLDYQLGNSSHP-----AVNMHLGSLLETGDGSG 120  
Db 59 QRSLYOC-----DYSTG-SCEPIRLQVPEAVNMSLGLSLAATTSPQ 101  
Qy 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQR-CPTY-MDVVIVLDGNSI 178  
Db 102 LLACGPTVHQTCSNTVVKGLFCFLGSLNQPPQPKFFEARLRCQPDSDIAFLIDGSGSI 161  
Qy 179 YPNSVOTFRLRVGKLFIDP--EQIO-----VGLVOYGESPVHNSLGDPTKEEVVRA 231  
Db 162 IPHD-----FRFM--KEFVSTWMEQLKSKTFLFSLMQYSSEEFRIHFKFQNNPRSL 214  
Qy 232 AKNLSREGRKTTAQAIMVACTGEGFSQSHGGRPEAARLLVVVTDGSHDGEELPAALKA 291  
Db 215 VKPITQLLGR--THTATGIRKVRRELFNITGARKNAFKILVITDGEKF-GDPLGYEDVI 272  
Qy 292 CEAGR--VTRYGIAVLGHYLRQRDPSSFLREIRTIASDPDERFFNVTDSSAALTVIDA 349  
Db 273 PEADREGVIRYIGV-GDAFRSEKS-----RQELNTIASKPRDRHVQVNNFEALKTQNQ 327  
Qy 350 LGDRIFGLEGSHAENESSFGLEMSIQGFSTHRLKDGILFGMVGAYDGGSVLMLEGGHRL 409  
Db 328 LREKIFAIEGTQTGSSSSFEHMSQEGFSAITNSGPLLSTVGSYDWAGGVFLYTSREKS 387  
Qy 410 FPPMALEDEFPFPAALQNHAAVLYGVSVMLLRGRRLLFLSGAPFRHGRKVIAQLKKDG 469

[illegible]

RESULT 9  
RWHU1C

cell surface glycoprotein CD11c precursor - human

N; Alternate names: leukocyte

C;Species: Homo sapiens (man)  
C;Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text change 09-Jul-2004

C;Accession: A36584; A35543; S00864

R;Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.

J. Biol. Chem. 265, 12750-12751, 1990

A;Reference number: A36584  
A;Contenta: erratum

A; Collection: ETACUM  
A; Accession: A36584

A;Molecule type: DNA

A;Residues: 1-1163 &lt;COR&gt;

**A;Cross-references:** UNIPROT:P20702

R;Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.

J. Biol. Chem. 265, 2782-2788, 1990

A:Title: Genomic structure of an integrin alpha subunit, the leukocyte p150,95 molecule.  
A:Reference number: A35543; MUID:90153906; PMID:2303426  
A:Accession: A35543  
A:Molecule type: DNA  
A:Residues: 1-834 <CO2>  
A:Note: This sequence has been revised in reference A36584  
R:Corbi, A.L.; Miller, L.J.; O'Connor, K.; Larson, R.S.; Springer, T.A.  
EMBO J. 6, 4023-4028, 1987  
A:Title: cDNA cloning and complete primary structure of the alpha subunit of a leukocyte  
A:Reference number: S00864; MUID:88166645; PMID:3327657

Query Match	17.9%	Score 1062.5;	DB 1;	Length 1163;
Best Local Similarity	28.4%	Pred. No. 2e-70;		
Matches 347;	Conservative 201;	Mismatches 476;	Indels 199;	Gaps 47;

J. Biol. Chem. 265, 2782-2788, 1990







A;Title: Aberrant upregulation of a novel integrin alpha subunit gene at 3p21.3 in small  
A;Reference number: 158409; MUID:94119603; PMID:8290272  
A;Accession: 158409  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-1035 <RES>  
A;Cross-references: UNIPROT:Q13797; GB:D25303; NID:9464180; PIDN:BA04984.1; PID:g533327  
R;Palmer, S.L.; Ruegg, C.; Ferrando, R.; Pytel, R.; Sheppard, D.  
J. Cell Biol. 123, 1289-1297, 1993  
A;Title: Sequence and tissue distribution of the integrin alpha9 subunit, a novel partne  
A;Reference number: A49459; MUID:94064789; PMID:8245132  
A;Accession: A49459  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 30-1035 <PAL>  
A;Cross-references: GB:I24158  
C;Superfamily: integrin alpha-4 chain  
C;Keywords: glycoprotein; metal binding; transmembrane protein  
F;1-27/Domain: signal sequence #status predicted <SIG>  
  
Query Match 12.0%; Score 714.5; DB 2; Length 1035;  
Best Local Similarity 22.4%; Pred. No. 1.5e-44;  
Matches 280; Conservative 170; Mismatches 432; Indels 369; Gaps 43;  
  
Qy 9 LFLPLVELTGLCSFNLDEHHPRLFPQPPPEAEFGYSVLQVGGQRMVLCAPWD----G 64  
Db 16 LLLALVAGIPAGAYNLDPQRPVHFQGPADSFYGVLEHFDNTRVWLVGAPKADSKYS 75  
Qy 65 PSGDRRGDVPCVGGAHNAP-----CAKHLGDYQLGNS-SHPANMHLQMSLL-E 114  
Db 76 PSVKSPGAVFKRV---HTNDRCTELDMARGNRTGSCOKTCEBDDDMGVSLARQ 132  
Qy 115 TDGSGFMACAPLWSRAC--GSSVFSSGICARVDASFQPGSLAPTAQRCPTVMYVIVL 172  
Db 133 PKADGRVLACAHKWNIVYEADHILPHGFCVLIIPSNLQAKG----- 173  
Qy 173 DGSNIYPSWSEVQTLRLVLGKLFIDPEIQVGLVQGESPVHWSLGDFTKKEVVRAA 232  
Db 174 -----RT----- 175  
Qy 233 KNLRSREGRETQTAIWVACTEGFSQSHGGRPEAARLLVVVTDGESHGDELPALAKAC 292  
Db 176 -----LIPCEYKKY-----GEEHG-----SC 194  
Qy 293 EAGRWTRYGIAVLGHLRRQDPSPFLREIRTIASDPDERFFNVNTDEAALTDIVDALGD 352  
Db 195 QAG-----IA----- 199  
Qy 353 RIFLGSHSAENSSFGLEMSQIGFSTRHLKDGILFGMVGAYDNGGVSVMLEGGHRLPPP 412  
Db 200 -----GFTBEL---VVMGAPGSFYWAGTIKVLNLTDNTY-- 231  
Qy 413 RMALEDPPPALQNHAAVLGYSVSSMLLGRGRRL-FLSGAPFRHRHGKVIQFOL-KKQGA 470  
Db 232 -LKUNDEV--IMNRYTYLGVAVTAGHFSHSTTDVVGAGQDGKIGKVIYIFRADRSQT 288  
Qy 471 VRVAQSLQGEIQIGYFSGELCPDLTRDGTGTDVLLVAAPMFLGPONKETGRVYVLYGQQ 530  
Db 289 LKIFQAGSKGWSYFGSSLCVLDLNGSLD-LLVGAPMF--SEIRDEGQVTVINRGN 345  
Qy 531 SLLTLQGTLPPEPQDARFGFAMGALPDNLQDGFADVAVGAPLEDGHOGALYLYHGTQSG 590  
Db 346 GALBEQALATGDGAYNAHFGEISLASLDLDNDGFPDVAIGAPKEDDFAGAYIYHGDAGG 405  
Qy 591 VRHPAQRIAAASMPHALSYGRSVDGLDLDGDDLDVAVGA--QGAAILSSRPVHL 648  
Db 406 IPVQYSMKLSGQKINPVLRMFGQSSIGGIDMDNGYPDVTVGAFMSDSVLLRLARPVITV 465  
Qy 649 TPLSEVTPQATISVQVQDCRRRQGEAVCLTAALCF-----QVTSRTP 689  
Db 466 DVSI-FLPGSINITAPQCHDQOQPVNCLNVTTCFSFHGKHVPEEIGLNYVLMADVAKKEK 524  
Qy 690 GRWDHQFYMRFTASLDEWTAGARAAFDGSGQRLLSPRRLRLSVGNVTCQLHFIHV-LDTS 748

Db 525 GQMPRVYFVLLGSTMGOVT-----EKLQTYMEETCRHVVAHVRRVQD 568  
Qy 749 YLRPVALVTVFAL-----DNTTKP-GPVL--NEGSPTSIQKLVPFSCDGPNECVT 797  
Db 569 VISPIVFEAAYSLSSEHVTGSEERELPELTPVLRAWKKGOKIAQKQTVFERNCRSE-DCAA 627  
Qy 798 DLVLQVNDIRG-SRKAPFVVRGRRKVLVSTLTLENKENAYNTLSLIIISRNHLASLT 856  
Db 628 DLQLOGLKLLSSMDEKTYLALGAVKNISLNISSNLGDDAYDANVSNFVNSRELFNNMW 687  
Qy 857 PORESPTKVECAAPSAHARLCSVGHVPVOTGAKVTFLLEFEFSCSSLLSOVFGKL-TASS 915  
Db 688 QKEE--MGISCLELSEDFKCSVGFPFMRSKSYKFSVIFDTSHLSGEEVLSFIVTAQS 745  
Qy 916 DSLRNGTLOENTAQTSAYIQE-----PHLLFSSSES----- 947  
Db 746 GNTSESLDNTLVLMVPLMHEVDTSITGMSPTSPVYGESVDAAANFIQLDDLECHPOP 805  
Qy 948 ---TLHRYEHPYGTLP-VGPGPEFKTTLTNNNA-----SCIVQNLTPEPP 988  
Db 806 INITLQVYNTGP-STLPGSSVSISFPNRLSSGGAEHFVQEMVVGQKNGCSFQKNPTPC 864  
Qy 989 GPPVHPELOHT-----NRLNGSNTQCVVRCHLGQAKGTEVSVGLLRLVHN 1036  
Db 865 IIFQEQENIFHTIPAFFTKSGRKVLDCEKPGISCLTAHCFSAKAKESRTIDIYMLNT 924  
Qy 1037 E-----PFRRAKFK---SLTVVSTFELGTEGSLVQLTEASRWSSESLLEVQTR 1082  
Db 925 EILKDDSSSIQFMSRAKVDPALRVVEIAHGNEPVTV--VFEA-----LHNLEPR 975  
Qy 1083 PILISLWILGSLVGLLLALLVFLCWLKGLFP--AHKKIPEEKREKLE 1131  
Db 976 GYVVG-WIAISLVGLILFLLAVLLWKGFFRRYKEIEAEKNRKE 1025  
  
RESULT 13  
JC7294  
alpha integrin - sea urchin (Strongylocentrotus purpuratus)  
C;Species: Strongylocentrotus purpuratus (purple urchin)  
C;Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 09-Jul-2004  
C;Accession: JC7294  
R;Susan, J.M.; Just, M.L.; Lennarz, W.J.  
Biochem. Biophys. Res. Commun. 272, 929-935, 2000  
A;Title: Cloning and characterization of alphaP integrin in embryos of the sea urchin St  
A;Reference number: JC7294  
A;Contents: Embryo  
A;Accession: JC7294  
A;Molecule type: mRNA  
A;Residues: 1-1054 <SUS>  
A;Cross-references: UNIPROT:Q9U6S1; GB:AAD55724  
C;Genetics:  
C;Gene: sualphap  
C;Superfamily: integrin alpha-2b chain  
C;Keywords: calcium binding; embryo; glycolysis; glycoprotein; heterodimer; transmembran  
  
Query Match 11.8%; Score 700; DB 2; Length 1054;  
Best Local Similarity 22.7%; Pred. No. 1.9e-43;  
Matches 286; Conservative 182; Mismatches 412; Indels 378; Gaps 45;  
  
Qy 10 FLPLVFLITGL-----CSSFNLDHHPRLFPQPPPEAEFGYSVLQVGGQRMVLYG---AP 61  
Db 4 FLLLSFCTTLLVLLDSTVAFNFDLRAPYKFDGPFQSGSLFGFVAQHRDQNTDVLVIGAPEP 63  
Qy 62 WDGPSSGRDRGDVRCYVGG-AHNAPCAK---GHLGDYQLGNSHSPAVNMHLGMSLLETDG 117  
Db 64 TTQGVTVNGAVYKCPVTPLPSGSCFCEQVPTDTTGTNEVLDKS-----NQWFGATLASSGP 119  
Qy 118 DGGFMACAP--LWSRACGSSVSFSSGICARVDASFQPGSLAPTAQRCPTVMYVIVLDGS 175  
Db 120 DGRILACAPRLVWLQT-----STISPT----- 141  
Qy 176 NSIYPSWSEVQTLRLVLGKLFIDPEIQVGLVQGESPVHWSLGDFTKKEVVRAAKNL 235

Db 142 ----- 141  
Qy 236 SRREGRTKTAQAIWVACTBGFSGHGRPEAARLLVVVTDGSHDGEELPAALKACEAG 295  
Db 142 -----DKREPT----- 148  
Qy 296 RVTRYGIAVLGHVLRQRDPSSFLREIIRTIASDPERFFNFVTDAAALTDIVDALGRIF 355  
Db 149 -----GTCFVG-----SDFTNFVNSPCQSTDRD-----LF 175  
Qy 356 GLGSHABNESSGLESQIGFSTHRLKDG--LLFGMVGYDMGGSVLMLEGGHRLPPPR 413  
Db 176 GFD-----KITHCOAGFSAQIPSDNSTLMGAPGSY-----YLQG--QIPAQ 216  
Qy 414 MA-LED-----EPPALQNAHAAYLGYSVSSMLLRG--GRRFLGAPRFRH--RGKVIAPQL 465  
Db 217 LSTLSVNSNTEQNAVFDN--SYNGISLALGDFNGDGLDYVVGTPRGESLRLGLVAFD- 273  
Qy 466 KKGAVRAVQSLQGEQIGSYFGSELCLPDTDRDGTDLVLLVAAPMFL-----GPNKETGR 521  
Db 274 --QSLVEIITPVVGEQIVSYFGYSVASVDVNGDGLDD--LLVGAPMFTNREPATEKWEAGR 330  
Qy 522 VYVYL-----VGOQSLLTLOGLTQPEPPQDARFGFANGALPDNLQDGFADVAVGABLE- 574  
Db 331 VYVYLNADHSLGAPOMLTGKKI-----RAREGFPIITSIGDSNQDGFNDVAIGAPYDG 383  
Qy 575 DGHGALVLYHGTGSGVRPHPAQRIAAASMPHA--LSYFGRSVDGRLLDGDGLVDVAVGA 633  
Db 384 EDNSGVYIITHGSAEGLRLTESQVLTPELGFSDIITTFGFSVDGQDQNDPDLVVGA 443  
Qy 634 QG--AAILLSRPIVHLTPSLVTPQAISSVVQRDCR--RRQGEAVCLTAALCFQVT--SRTP 689  
Db 444 ESADAALVTRPVVLLAEALTIEPIGLNDKNKYELPDGTWTSFVAMACFTYTGHNLP 503  
Qy 690 GRWDHQYPMFTASLDWTAGARAAPDGSQORLSPRRLRLSVGNVTCEQLHFHVLDT--SD 748  
Db 504 ARIGISYTLTVDDSI-----TSGRRALLEVNLSQVTKRNLVDGMKFCOPLRAYVYNTIQD 560  
Qy 749 YLRPVALTVTFALDNTT-----KPGPVNLGSGPSTSIQKLVPPSKDGPDCNECVTDLVLQV 803  
Db 561 KLTPVAVDQYELTDESILPYELPIINKEAVSSQTKVSIQNNC--VNNICPIEIGITV 619  
Qy 804 NMDIRGSKAPFVVRGRRKRVVSTTLLENKENAYNTSLIIFSRNLHLASLTQPRESPI 863  
Db 620 TPNL-----PNVIGQAQELTLVVSINNRGEDAFQSTLAVYYPEGLQYVRLERRANWDF 673  
Qy 864 KVECAAPSAHARL--CSVGHPVFQGAQVTFLEPEFSCSLLSQVFK-----LTASS 915  
Db 674 SVTCTEDSALRMITCDTGNPLVG-----KYNLEFGLTLSTL--QVSGDKONIEFYLVA 726  
Qy 916 DSLERNGTLOENTAQTSAIYQEPHLLPSSSES----- 947  
Db 727 ENNEPDNTLONNELNVTAIVVDATLKLSSASYPEIVTVRVPEDNIVPEPPTKNASEADI 786  
Qy 948 ---TLHRYEHPYG-----TL-----P 961  
Db 787 GMEVVHLYEVRNTGSSNAAEVTLNIRWPKDENGDFLYLLGIMTDSGVTQISQGANP 846  
Qy 962 VGPOPEFKTLTRTNNAACIVQNLTEPPGPPVHPPELQHTNRLNGSNTOCQVVRCHLQOLA 1021  
Db 847 LGVKLEASTKEQLSNSTVSGRKRKEG--YAEALQAQBPFICTPESCVLINCTIDEIK 904  
Qy 1022 KGTBVSGLLRVHNEPFRAPKSLTVVSTFELGTSEGSV-----LQLTEASRW 1071  
Db 905 ATKSVVRIILGRFWERTFQKAVSAVPIQVTLASTATATVRSIPYNIPIPMETDSTKA 964  
Qy 1072 SESL--LEVVOVTRPILISLWILIGSVLGLLALLVFLCMLKGLGFAHKKIPPEEKREE 1128  
Db 965 STLITABELVLPVVSIAWIIIVSVVLGGIILLIIILGLMKCGFFERKKPGEDQKEYE 1022

RESULT 14

A41131

lymphocyte-Peyer's patch adhesion molecule alpha 4 precursor - mouse

N;Alternate names: integrin alpha-4

C;Species: Mus musculus (house mouse)

C;Date: 05-Jun-1992 #sequence\_revision 05-Jun-1992 #text\_change 09-Jul-2004

C;Accession: A41131; S16742

R;Neuhaus, H.; Hu, M.C.T.; Hemler, M.E.; Takada, Y.; Holzmann, B.; Weissman, I.L.

J. Cell Biol. 115, 1149-1158, 1991

A;Title: Cloning and expression of cDNAs for the alpha subunit of the murine lymphocyte-

A;Reference number: A41131; MUID:92064645; PMID:1840602

A;Accession: A41131

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1039 &lt;NEU&gt;

A;Cross-references: UNIPROT:Q00651; EMBL:X53176; NID:g51484; PIDN:CAA37316.1; PID:g51485

C;Superfamily: integrin alpha-4 chain

C;Keywords: cytoskeleton; transmembrane protein

Query Match 11.2%; Score 666.5; DB 2; Length 1039;  
Best Local Similarity 22.4%; Pred. No. 5.7e-41;  
Matches 279; Conservative 184; Mismatches 382; Indels 403; Gaps 58;

Qy 23 FNLDEHHRLPFCPPPEAREFGYSVLQHVGGQRMVLVGP---W-DGPGDRRGDVVRCPV 78

Db 41 YNLDPENALLYQPGSGTLFGYSVYLHSHGSKRWLIIVGAPTASWLSNASVVPNGAIYRCGI 100

Qy 79 -----GGAHNAPCAKGLHDYQLGNSHSPAVNMHGLMSLLETDGD--GGFMAC 124

Db 101 RKNPNQTCBQLQSGSPGPGCKTCLBERD-----NQMLGVTLRSQPGNGSIVTC 151

Qy 125 APLWSRACGSSVSSFGICARVDASFPQGSLSAPTAQRCPTMYDVIIVLDGNSIYPMSEV 184

Db 152 GHRW-----KNIP-----YMKSDNKLPT----- 169

Qy 185 OTFLRLVGLFLDPBQIQVGLVOYGESPVHEWSLSDPRTKEVVRRAAKNLSRREGRETK 244

Db 170 -----GICYMP-----SDLRTE-----LSKR----- 186

Qy 245 TAQAIWVACTBGFSGHGRPEAARLLVVVTDGSHDGEELPAALKACEAGRVTRYGI 304

Db 187 -----MAPCYKDYTRFEGN-----PASCQAG----- 208

Qy 305 LGHYLRQRDPSPFLREIRTIASDPDER-----PFFNVTEAALTDIVDALGRIFGLE 358

Db 209 -----ISSFYQDLIVMGAPGSSYMTGVTVVNIIT-----TNOYKAFVDR----- 248

Qy 359 GSHABNESSGLEMSQIGFSTHRLKDGILLFGMVGYDMGGSVLMLEGGHRLFPFMALED 418

Db 249 -----QNOVKFG----- 255

Qy 419 EPPALQNAHAAYLGYSVSSMLLRGGRRLL--FLSGAPRFRHRGKVIAFQKKDGAVRVAQSL 477

Db 256 -----SYLGYSVGAGHFRSPHTTEVVGAPQHBQIGKAYIYSI--DENELNIYEM 304

Qy 478 QGEQIGSYFGSELCLPDTDRDGTDLVLLVAAPMFLGPQN--KETGRVYVYLVGOQS--LL 533

Db 305 KGKGLSGSYFGASCAVDNLADGFS--LLVGPW---QSTIREGRVFFVYVINGMGAVMV 359

Qy 534 TLOGTLQPEPPQDARFGFANGALPDNLQDGFADVAVGAPLEDHQGALYLYHGTQSGVRP 593

Db 360 EMERVLVSGDKYAAARFGESIANGLDNDGFDIATGAPQEDDLRGAVIYNGRVDSISS 419

Qy 594 HPQRATAAAMPALSYFGRSVDGRLLDGDGLVDVAVGA--QQAAILLSRPIVHLTPS 651

Db 420 TYSQREGGQISKSLRMFGQISGQIDANNGYVDVAVGAFQSDSAVLLTRPVVIVEAS 479

Qy 652 LEVTPQAISSVVQDCCRRRQGEAVCLTAALCFQVTSR--TPGRWDHQYPMFRFTASLDE--- 706

Db 480 LS-HPESVNRKTCKDCTENGLPSVCMHLTLCFSYKGEVPG-----YIVLVFNVSILDVHRKA 534

Qy 707 -----WTAGARAAAFDGSQORLSPRRLRLSVGNVTCEQLH--FHVLDTSVLRPVALT 756

Db 535 ESPSRFYFFSNGTSDVITGS-----IRVSSGSKC-RTHQAFMRKQVRDILTPIHVE 585







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 167.54 Seconds  
(without alignments)  
3459.921 Million cell updates/sec

Title: US-09-647-544-4  
Perfect score: 5912  
Sequence: 1 MEKPFVTHFLPLVFLTGLC.....GFFAHKKIPEEKKEKLEQ 1132

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : UniProt 03.\*

1: uniprot\_sprot.\*  
2: uniprot\_crembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5904.5	99.5	1167	1 ITAG HUMAN	O75578 homo sapien
2	2344.5	39.5	1189	1 ITAH HUMAN	Q9ukx5 homo sapien
3	2333.5	39.3	1188	1 ITAH MOUSE	P61622 mus musculus
4	2333.5	39.3	1188	2 Q7TQC3	Q7tqc3 mus musculus
5	1884	31.8	1180	1 ITAL RAT	P18614 rattus norv
6	1881.5	31.7	1151	1 ITAL HUMAN	P56199 homo sapien
7	1827	30.8	1171	2 O42094	O42094 gallus gall
8	1734.5	29.2	1170	1 ITA2 BOVIN	P53710 bos taurus
9	1733.5	29.2	1181	1 ITA2 HUMAN	P17301 homo sapien
10	1712.5	28.9	1178	1 ITA2 MOUSE	Q62469 mus musculus
11	1712	28.9	1178	2 Q6P1C7	Q6p1c7 mus musculus
12	1403	23.7	823	2 Q8WY18	Q8wy18 homo sapien
13	1398.5	23.4	823	2 Q8CE94	Q8ce94 mus musculus
14	1135.5	19.1	1161	1 ITAD RAT	Q9eyr7 rattus norv
15	1126	19.0	1162	1 ITAD HUMAN	Q13349 homo sapien
16	1115.5	18.8	1170	1 ITAL HUMAN	P20701 homo sapien
17	1105.5	18.6	1163	1 ITAL MOUSE	P24063 mus musculus
18	1104.5	18.6	1161	2 Q9WTV4	Q9wtv4 mus musculus
19	1097	18.5	1160	2 Q9R200	Q9r200 mus musculus
20	1090	18.4	1188	2 Q8KAS4	Q6kas4 mus musculus
21	1084.5	18.3	1152	1 ITAM HUMAN	P11215 homo sapien
22	1077.5	18.2	288	2 Q8BM12	Q8bm12 mus musculus
23	1073.5	18.1	1169	1 ITAX MOUSE	Q9gmh4 mus musculus
24	1064	17.9	1165	1 ITAL BOVIN	P16255 bos taurus
25	1062.5	17.9	1163	1 ITAX HUMAN	P20702 homo sapien
26	1037	17.5	1166	2 Q6TYB8	Q6tyb8 bos taurus
27	1025	17.3	1153	1 ITAM MOUSE	P05555 mus musculus
28	1013	17.1	1151	2 Q9J130	Q9j130 rattus norv
29	994.5	16.8	1086	2 Q96HB1	Q96hb1 homo sapien
30	972.5	16.4	780	2 Q06271	Q06271 xenopus lae
31	972	16.4	1196	2 Q98TF1	Q98tf1 cyprinus ca

32	956.5	16.1	1167	1 ITAE MOUSE	O60677 mus musculus
33	956.5	16.1	1167	2 O88340	O88340 rattus norv
34	947.5	16.0	1167	2 O88341	O88341 rattus norv
35	937	15.8	1187	2 Q98TF0	P38570 cyprinus ca
36	928	15.6	1179	1 ITAE HUMAN	P38570 homo sapien
37	893	15.1	920	2 Q28984	Q28984 sus scrofa
38	891.5	15.0	1038	2 Q8BS01	O8be01 mus musculus
39	883	14.9	895	2 Q9WUF8	Q9wuf8 mus sp. itg
40	873.5	14.7	927	2 Q8HZV0	Q8hzv0 bos taurus
41	792	13.4	1160	2 Q8MKF4	Q8mkf4 felis silve
42	780.5	13.2	1332	2 Q9BPQ8	Q9bpq8 halocynthia
43	754.5	12.7	1036	2 Q91YD5	Q91yd5 mus musculus
44	714.5	12.0	1035	1 ITA9 HUMAN	Q13797 homo sapien
45	704.5	11.9	1033	2 Q9BGU3	Q9bg93 bos taurus

#### ALIGNMENTS

#### RESULT 1

ITAG\_HUMAN ID ITAG\_HUMAN STANDARD; PRT; 1167 AA.  
AC O75578; Q9UH28;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Integrin alpha-10 precursor.  
GN Name=ITGA10;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Articular chondrocytes;  
RC MEDLINE=98352078; PubMed=9685391; DOI=10.1074/jbc.273.32.20383;  
RA Camper L., Hellman U., Lundgren-Aakerlund E.;  
RT "Isolation, cloning, and sequence analysis of the integrin subunit  
RT alpha10, a beta1-associated collagen binding integrin expressed on  
RT chondrocytes.";  
RL J. Biol. Chem. 273:20383-20389(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Endothelial cells, and Heart;  
RX MEDLINE=20169197; PubMed=10702680;  
RA Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,  
RT Wang S.-X., Langley R., Krissansen G.W.;  
RT "The integrin alpha10 subunit: expression pattern, partial gene  
RT structure, and chromosomal localization.";  
RT Cytogenet. Cell Genet. 87:238-244(1999).  
CC -!- FUNCTION: Integrin alpha-10/beta-1 is a receptor for collagen.  
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-10  
CC associates with beta-1.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- TISSUE SPECIFICITY: Widely expressed with highest expression in  
CC muscle and heart. Found in articular cartilage.  
CC -!- DOMAIN: The integrin I-domain (insert) is a VWF domain. Integrins  
CC with I-domains do not undergo protease cleavage.  
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.  
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.  
CC -!- SIMILARITY: Contains 1 VWF domain.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----

EMBL; AF074015; AAC31952.1; --  
EMBL; AF112345; AAF21944.1; --  
EMBL; AF172723; AAF61638.1; --

DR HSP; P18614; 1MHP.  
DR Genew; HGNC:6135; ITGA10.  
DR MIM; 604042; -.  
DR GO; GO:0008305; C:integrin complex; TAS.  
DR GO; GO:0005518; F:collagen binding; TAS.  
DR GO; GO:0007160; F:cell-matrix adhesion; TAS.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF01839; FG-GAP; 3.  
DR Pfam; PF00092; VWA; 1.  
DR PRINTS; PRO1185; INTEGRINA.  
DR PRINTS; PRO0453; VWFADOMAIN.  
DR SMART; SM00191; Int\_alpha; 5.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS00242; INTEGRIN\_ALPHA; FALSE\_NEG.  
DR PROSITE; PS02034; VWA; 1.  
DR Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;  
KW Repeat; Signal; Transmembrane.  
FT SIGNAL 1 22 Potential.  
FT CHAIN 23 1167 Integrin alpha-10.  
FT DOMAIN 23 1122 Extracellular (Potential).  
FT TRANSMEM 1123 1145 Potential.  
FT DOMAIN 1146 1167 Cytoplasmic (Potential).  
FT REPEAT 38 97 FG-GAP 1.  
FT REPEAT 98 ? FG-GAP 2.  
FT DOMAIN 167 350 VWA.  
FT REPEAT 365 427 FG-GAP 3.  
FT REPEAT 428 482 FG-GAP 4.  
FT REPEAT 483 545 FG-GAP 5.  
FT REPEAT 546 605 FG-GAP 6.  
FT REPEAT 608 660 FG-GAP 7.  
FT DOMAIN 1134 1140 Poly-Leu.  
FT CA\_BIND 494 502 Potential.  
FT CA\_BIND 558 566 Potential.  
FT CA\_BIND 620 628 Potential.  
FT DISULFID 76 86 By similarity.  
FT DISULFID 666 675 By similarity.  
FT DISULFID 681 736 By similarity.  
FT DISULFID 789 795 By similarity.  
FT CARBOHYD 98 98 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 234 234 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 336 336 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 364 364 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 733 733 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 763 763 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 839 839 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 921 921 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 1011 1011 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 1018 1018 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 1039 1039 N-linked (GlcNAc... ) (Potential).  
FT CONFLICT 844 844 I -> L (in Ref. 2).  
FT CONFLICT 909 909 G -> V (in Ref. 2).  
FT CONFLICT 926 926 E -> D (in Ref. 2).  
SQ SEQUENCE 1167 AA; 127573 MW; AE7D3A1C25C1AB0 CRC64;

Query Match 99.5%; Score 5904.5; DB 1; Length 1167;  
Best Local Similarity 97.0%; Pred. No. 0;  
Matches 1132; Conservative 0; Mismatches 0; Indels 35; Gaps 1;

Qy 1 MELPFVTHLPVFLVLTGLCSPPNLDHHPRLFPQPEAEFGYSLQHVGGQRMVLVGA 60  
Dy 1 MELPFVTHLPVFLVLTGLCSPPNLDHHPRLFPQPEAEFGYSLQHVGGQRMVLVGA 60  
Qy 61 PWDGSGDRGDRVRCVGGAAHAPCAKGLHGDYQLGNSSHPAVNMHLGMSLLTDDGG 120  
Dy 61 PWDGSGDRGDRVRCVGGAAHAPCAKGLHGDYQLGNSSHPAVNMHLGMSLLTDDGG 120  
Qy 121 FMACAPLWSRACGSGSVFSSGICARVDASFPQGSGLAPTAQRCPTYMDVIVLGDGNSIYP 180  
Dy 121 FMACAPLWSRACGSGSVFSSGICARVDASFPQGSGLAPTAQRCPTYMDVIVLGDGNSIYP 180  
Qy 181 WSEVQTFLLRLVGLKFLDIDPEIQVGLVQYGSPVHWSLGDFTKBEVVRAAKNLSRREG 240

181 WSEVQTFLLRLVGLKFLDIDPEIQVGLVQYGSPVHWSLGDFTKBEVVRAAKNLSRREG 240  
241 RETKTAQIMVACTEGFSQSHGGRPEAARLLVVVTGESHGDBELPAALKACBAGRVTY 300  
241 RETKTAQIMVACTEGFSQSHGGRPEAARLLVVVTGESHGDBELPAALKACBAGRVTY 300  
301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNVTVDEAALTDIVDALGDRIFGLEGS 360  
301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNVTVDEAALTDIVDALGDRIFGLEGS 360  
361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGAYDWGSGVLMLEGHRLPPRMALEDEF 420  
361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGAYDWGSGVLMLEGHRLPPRMALEDEF 420  
421 PPALQNHAAVLGYSVSMILRGRRLLFLSGAPFRHRGKVIAFOLKDKGAVRVAQSLQGE 480  
421 PPALQNHAAVLGYSVSMILRGRRLLFLSGAPFRHRGKVIAFOLKDKGAVRVAQSLQGE 480  
481 QIGSYFGSELCPDTRDGTDDVLLVAAMPFLGPQNKETGRVYVYLVGQOSLLTLQSTLQ 540  
481 QIGSYFGSELCPDTRDGTDDVLLVAAMPFLGPQNKETGRVYVYLVGQOSLLTLQSTLQ 540  
541 PEPPQDARFGFAMGALPDNLNODGFADVAVGAPLEDHGHGALYLYHGTQSGVRPHPAQRIA 600  
541 PEPPQDARFGFAMGALPDNLNODGFADVAVGAPLEDHGHGALYLYHGTQSGVRPHPAQRIA 600  
601 AASMPHALSYFGRSVDGRLLDGDLDVDAVGAQGAAILLSRPIVHLTPSLEVTPOAIS 660  
601 AASMPHALSYFGRSVDGRLLDGDLDVDAVGAQGAAILLSRPIVHLTPSLEVTPOAIS 660  
661 VVQDRCRRQGEAVCLTAALCFQVTSRTPGRWDHQFYMRFTASLDEWTAGARAAFDGSGQ 720  
661 VVQDRCRRQGEAVCLTAALCFQVTSRTPGRWDHQFYMRFTASLDEWTAGARAAFDGSGQ 720  
721 RLSPRRLRLSVGNVTCBQLHFHVLDTSDYLRPVALTVPFALDNTTKPGPVLNEGSPTSIQ 780  
721 RLSPRRLRLSVGNVTCBQLHFHVLDTSDYLRPVALTVPFALDNTTKPGPVLNEGSPTSIQ 780  
781 KLVPPSKDCGPDNECVTDLVQNMDIRGSRKAPFVVRGRRKVLVSTTLLENRKENAYNT 840  
781 KLVPPSKDCGPDNECVTDLVQNMDIRGSRKAPFVVRGRRKVLVSTTLLENRKENAYNT 840  
841 SLSIIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPVFTGAKVTFLEPFESC 900  
841 SLSIIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPVFTGAKVTFLEPFESC 900  
901 SLLSQVFGKLTASSDSLENGTLQENTATSAIYIEPHLLFSSESILHRYEYVHPYGTLL 960  
901 SLLSQVFGKLTASSDSLENGTLQENTATSAIYIEPHLLFSSESILHRYEYVHPYGTLL 960  
961 PVGPGPEFKTTLR-----TNNASCIVQNL 985  
961 PVGPGPEFKTTLR-----TNNASCIVQNL 985  
986 EPPGPPVHPELOHTNRNGSNSTQCQVVRCHLQAKGTEVSQVGLLRVHNEFFERRAKFK 1045  
986 EPPGPPVHPELOHTNRNGSNSTQCQVVRCHLQAKGTEVSQVGLLRVHNEFFERRAKFK 1045  
1021 EPPGPPVHPELOHTNRNGSNSTQCQVVRCHLQAKGTEVSQVGLLRVHNEFFERRAKFK 1080  
1046 SLTVVSTFELGTREGSVLQTEASRWSESLELVQVTPRILISWILGSLVGLGLLALL 1105  
1081 SLTVVSTFELGTREGSVLQTEASRWSESLELVQVTPRILISWILGSLVGLGLLALL 1140  
1106 VFCLMKLGFFAHKKIPEEKREKLEQ 1132  
1141 VFCLMKLGFFAHKKIPEEKREKLEQ 1167

RESULT 2  
ITAH HUMAN  
ID ITAH HUMAN STANDARD; PRY; 1189 AA.  
AC Q9UKX5; Q9UKQ1;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-11 precursor.
GN Name=ITGALL;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal heart, and Osteoblast;
RX MEDLINE=99417678; PubMed=1048209; DOI=10.1074/jbc.274.36.25735;
RA Velling T., Kusche-Gullberg M., Sejersen T., Gullberg D.;
RA Lehnert K., Ni J., Leung E., Gough S.M., Weaver A., Yao W.P., Liu D.,
RA Wang S.-X., Morris C.M., Kriksansen G.W.;
RT "Cloning, sequence analysis, and chromosomal localization of the novel
RT human integrin alpha11 subunit (ITGALL).";
RL Genomics 60:179-187(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal muscle, and Uterus;
RX MEDLINE=99395147; PubMed=10464311; DOI=10.1074/jbc.274.36.25735;
RA Velling T., Kusche-Gullberg M., Sejersen T., Gullberg D.;
RT "cDNA Cloning and Chromosomal Localization of Human alpha(11)
RT Integrin. A collagen-binding, i domain-containing, beta(1)-associated
RT integrin alpha-chain present in muscle tissues.";
RL J. Biol. Chem. 274:25735-25742(1999).
RN [3]
RP SEQUENCE OF 954-1188 FROM N.A.
RC TISSUE=Fibroblast;
RA Andreu N., Estivill X., Escarceller M., Sunoy L.;
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RL FUNCTION: Integrin alpha-11/beta-1 is a receptor for collagen.
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-11
CC associates with beta-1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: According Ref.2 highest levels in uterus and
CC heart, intermediate levels in skeletal muscle and intermediate to
CC low levels in pancreas, kidney and placenta. According to Ref.1
CC also found in brain, colon, lung, small intestine, stomach,
CC testis, salivary glands, thyroid glands and prostate. Very low
CC levels in peripheral blood lymphocytes, fetal brain and fetal
CC liver.
CC -!- DEVELOPMENTAL STAGE: Strongly up-regulated in differentiating
CC fetal muscle cells (in vitro).
CC -!- DOMAIN: The integrin i-domain (insert) is a VWFA domain. Integrins
CC with i-domains do not undergo protease cleavage.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF109681; AA01258.1; -
DR EMBL; AF137378; AA051919.2; -
DR EMBL; AL359064; CAB94392.1; -
DR HSSP; P18614; IMHP.
DR Genew; HGNC:6136; ITGALL.
DR MIN; 604789; -
DR GO; GO:0008305; C:integrin complex; TAS.
DR GO; GO:0005518; C:collagen binding; TAS.
DR GO; GO:0007160; P:cell-matrix adhesion; TAS.
DR GO; GO:0007517; P:muscle development; TAS.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00092; VWFA; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
```

```
DR SMART; SMO0191; Int_alpha; 5.
DR SMART; SMO0327; VWFA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
DR PROSITE; PS0234; VWFA; 1.
KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium;
KW Polymorphism; Receptor; Repeat; Signal; Transmembrane.
FT SIGNAL 1 22
FT CHAIN 23 1189
FT DOMAIN 23 1142
FT TRANSMEM 1143 1165
FT DOMAIN 1166 1189
FT REPEAT 38 94
FT REPEAT 102 163
FT DOMAIN 164 345
FT REPEAT 359 420
FT REPEAT 422 475
FT REPEAT 477 537
FT REPEAT 539 598
FT REPEAT 601 653
FT DOMAIN 1154 1162
FT DOMAIN 1174 1177
FT CA_BIND 488 496
FT CA_BIND 551 559
FT CA_BIND 613 621
FT DISULFID 76 83
FT DISULFID 121 139
FT DISULFID 129 159
FT DISULFID 659 668
FT DISULFID 674 729
FT DISULFID 781 787
FT DISULFID 881 893
FT CARBOHYD 82 82
FT CARBOHYD 95 95
FT CARBOHYD 291 291
FT CARBOHYD 331 331
FT CARBOHYD 358 358
FT CARBOHYD 449 449
FT CARBOHYD 462 462
FT CARBOHYD 528 528
FT CARBOHYD 642 642
FT CARBOHYD 694 694
FT CARBOHYD 857 857
FT CARBOHYD 894 894
FT CARBOHYD 973 973
FT CARBOHYD 1032 1032
FT CARBOHYD 1040 1040
FT VARIANT 433 433
FT VARIANT 524 524
FT VARIANT 972 972
FT VARIANT 1003 1003
FT VARIANT 1030 1030
FT VARIANT 1094 1094
FT SEQUENCE 1189 AA; 133609 MW; 60303C08A4A4CD52 CRC64;

Query Match 39.5%; Score 2344.5; DB 1; Length 1189;
Best Local Similarity 42.2%; Pred. No. 6.6e-159;
Matches 505; Conservative 204; Mismatches 407; Indels 81; Gaps 16;

Qy 1 MELPFTVTHLPLVFLTGLCSFPNLDHPRLPFGPPEAEFGVSLQHVGGQRMVLVGA 60
Db 1 MDLPGRGLVVAWALSLLWPGFTDTFMMDTRKRPVTPGSTAFPGYTVQQDHSIGNKMLVGA 60
Qy 61 PWDGPGSGDRRGDYRCPVGGAHNAPCAKGLHGLYQLGNSSHHPAVNMHLGMSLLETGDGG 120
Db 61 PLETNQYQKGTGDVYKCPV---IHGNC TKLNLGRVTLNLSERKDNMRGLSLATNPXONS 117
Qy 121 FMACAPLWSRACSSGVFSSGICARVDASFQPGQSLAFTAQRCTPYMDVIVLDGNSIYP 180
```

Db 118 FLACPLWSECGSYTTGCSRVNFRFSKVPALQRCQTYMDIVVLGDSNIYP 177  
Qy 181 WSEVOTLRLRLVGLKFLDPEQIOGLVOYGESPVHMSLGDFTKEVVRPAKNSLRREG 240  
Db 178 WVEVQHFLINLLKFFYIGQOIQGVVOYGEDVHHEFLNDYRSVKDVEAAASHIRGG 237  
Qy 241 RETTAQAIWACTEGFSQHGGRPEAARLLVVVTDGESHGDBELPAALKACAGRVTRY 300  
Db 238 TETRTAFGIFARSEAFQK--GGRKGAKVMIVITDGESHSDPLEKVIQOESRDNTRY 295  
Qy 301 GIAVLGHLRQRDPSPFLREIRTIASDPDERFFNVYDEAALTDIVDALGDRIFGLEGS 360  
Db 296 AVAVLGYNRRGINPETFLEIKYIASDPDKFNFVYDEAALKDIDVALGDRIFSLEGT 355  
Qy 361 HAENESFGLMSQIGFSTHRLKDGILFGWGVAGVDMGSGVLMLEGGHRLPPRMALEDEF 420  
Db 356 N-KNETSFGLEMSQTGFSSHVEDGVLLGAVGDMGAVLKETSAGKVIPLRESYLKBF 414  
Qy 421 PPALQNHAAVLGYSSVSMLLRGRRFLSGAPRFRHKGKVIATFOLKKDGVAVVAQSLQGE 480  
Db 415 PEELKNHGAIVLYTVTSVSSRQGRVYVAGAPRNFHTGKVLFTMHNRSILTIHQMRGQ 474  
Qy 481 QIGSYFSGELCPLDTRDGTDLVLLVAAPMLFGPQNKETGRVYVYLVGQSSLLTFLQTLQ 540  
Db 475 QIGSYFSGEITSVDIDGVDVLLVGAAPMYFN--EGREGRKVVYVEL--RQNRFYVNGTLK 532  
Qy 541 PEPP-QDARGCFANGALPDNLNOGFADVAVGAPLEDHOGALYLYHGTQSGVRPHPRQRI 599  
Db 533 DSHSYQNRFGSSIASVDRLNQDSYNDVVGAPLEDNHAGAIYIFHGFSGSILKTPQRI 592  
Qy 600 AAASMPHALSYFGRSVDRLDGLDDLDVAVAGAAIILSSRPVHLTPSLSEVTPQAI 659  
Db 593 TASELATGLQYFGCSIHQGLDLEDGLDIAVAGALGNVILWSPVQINASLHFESKI 652  
Qy 660 SVVQDCRRRQGEAVCLTAALCFQVTSRTPGWMDHQPYMFRTASLDWNTAGAAAFDGG 719  
Db 653 NIFHRDCKRSRQDCLAAFLCFTPIFLAPHFQTTTIGIRYNATMDERRYTTPRAHLDEGG 712  
Qy 720 QRLSPRLRLSVGNVTCQELHFLVDTSDYLRPVALTVPALDNTKPGPVNLGSPSTSI 779  
Db 713 DRFNRAVLSSGQELCEINFHVDYADYVKPFTFVSYSLED--GMLDDGWWTTL 771  
Qy 780 QKLVFSGKDCPDNECVTDLVQLVNMDI-----RGSRK-----APFV 817  
Db 772 RVSVFPWNGCNEDEHCVPLDVLARSDLPTAMEYQORVLRKPAQCSAYTILSPFTTVPIL 831  
Qy 818 RGGRRKVLVSTTLNRKENAYNTSLIFSRLHLASLTPORESPIKVECAAPS--AHAR 875  
Db 832 ESTQRVAVATLENRGENAYSTVLNISQSANLQFASLIQKEDSGSIECVNEERLQKQ 891  
Qy 876 LCSVGHVPFQTGAKVTLLEPERSCSLLSVOFGKLTASSDSLNRNGTLQENTATQTSAYI 935  
Db 892 VCNVSYFFFRKAKAVFRLDPEFSKIFLHHLELTAAGSDNSRDKEDNVAPLRFHL 951  
Qy 936 QYEPHLLFSSESTLHRYEVHPYGLP--VGPGEFKTLTRN----- 975  
Db 952 KYEADVLFTRSSLSHVEKLVNSSLERYVDGIGPPFCIFRQNLGLPFIHGMKTIPI 1011  
Qy 976 -----NASC-IVONLTPEPGPVHPPELQHTNRLNGSNTOCQVVR 1014  
Db 1012 ATRSGNRLKLRLDFTLDEVANTSCNIWGNSTERYPTPE--EDLRAPQLNHSNDVVSIN 1070  
Qy 1015 CHLQQLAGTQSVGLRLVHNEFRRAKPSLTVSTFELGTGEGSVLQLETSRWSSES 1074  
Db 1071 CNI--RLVNPQINFLHNLRLSLKALKYKSMKIMVNAALQROFHPFIFREDBPSRQI 1129  
Qy 1075 LLEVVTQRPILISWLIGSVLGGALLLALVFLKWLKGFPAHKIPKPEEKREKLE 1131  
Db 1130 VPEISKQEDWQVPIWIVGSTLGGLLALLLVALLKLGFRSAR----RRREFGLD 1182

ITAH MOUSE STANDARD; PRT; 1188 AA.  
AC P61622;  
DT 05-JUL-2004 (Rel. 44, Created)  
DT 05-JUL-2004 (Rel. 44, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Integrin alpha-11 precursor.  
GN Name=Itgall;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Klaunig R.D., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Pahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Kzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- FUNCTION: Integrin alpha-11/beta-1 is a receptor for collagen (By  
similarity).  
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-11  
associates with beta-1 (By similarity).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
CC -!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins  
with I-domains do not undergo protease cleavage.  
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.  
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.  
CC -!- SIMILARITY: Contains 1 VWFA domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.ebi.ac.uk/ebis-sib.ch>)  
or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).  
CC -----  
DR EMBL; BC058716; AAH58716.1; -;  
DR MGI; 2442114; Itgall.  
DR PROSITE; PS00242; INTEGRIN\_ALPHA; FALSE\_NEG.  
DR PROSITE; PS00234; VWFA; 1.  
KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;  
FT Repeat; Signal; Transmembrane.  
FT SIGNAL 1 22 Potential.  
FT CHAIN 23 1188 Integrin alpha-11.  
FT DOMAIN 23 1141 Extracellular (Potential).  
FT TRANSMEM 1142 1164 Potential.  
FT DOMAIN 1165 1188 Cytoplasmic.  
FT REPEAT 38 94 FG-GAP 1.  
FT REPEAT 102 163 FG-GAP 2.  
FT REPEAT 164 345 VWFA.  
FT DOMAIN 359 420 FG-GAP 3.  
FT REPEAT 422 475 FG-GAP 4.  
FT REPEAT 477 537 FG-GAP 5.  
FT REPEAT 539 598 FG-GAP 6.

FT	REPEAT	601	653	FG-CAP 7.	
FT	CA_BIND	488	496	Potential.	
FT	CA_BIND	551	559	Potential.	
FT	CA_BIND	613	621	Potential.	
FT	DISULFID	76	83	By similarity.	
FT	DISULFID	121	139	Potential.	
FT	DISULFID	129	159	Potential.	
FT	DISULFID	659	668	By similarity.	
FT	DISULFID	674	729	By similarity.	
FT	DISULFID	781	787	By similarity.	
FT	DISULFID	881	893	By similarity.	
FT	CARBOHYD	82	82	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	95	95	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	291	291	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	331	331	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	358	358	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	449	449	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	462	462	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	528	528	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	642	642	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	694	694	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	857	857	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	894	894	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	973	973	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	1031	1031	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	1039	1039	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	1059	1059	N-linked (GlcNAc. .)	(Potential).
FT	CARBOHYD	1188	1188	N-linked (GlcNAc. .)	(Potential).
FT	SEQUENCE	1188	1188	AA; 133011 MW; 0B0313C90D65422E CRC64;	
Query Match					39, 34; Score 2333.5; DB 1; Length 1188;
Best Local Similarity					42.24; Pred. No. 4e-158;
Matches					499; Conservative 200; Mismatches 398; Indels 85; Gaps 15;
Qy	18	GLCSFPNLDHPRLPFGPPEAFEGYSVLQVGGQRMVLMVAPMDGPGSDRRGDVYRCP	77		
Db	18	GFTDTFNMDTNPRVIAGPSAFAFGYTVQQHDISGKXWLVVAPMETNGHOKTGDDVYKCP	77		
Qy	78	VGGAHNAPCAKHGLDYLGNSSHPVNMHLCMSLLETDGDDGGMACAPLWSRACGSSVF	137		
Db	78	VTQGN---CTKLNLRVTLNVSRRKDNMRGLSLATNPKNDSFLACSLWSHCSSYY	134		
Qy	138	SSGICARVDASFOFGSLAPTAQRCPTVMDVVVLVDGNSIYPMSEVQTFLLRLVGLFI	197		
Db	135	TTGMSRANSFRSKVPALQRCQTMDIVLDGNSIYPMVEVQHFLINLKFFYI	194		
Qy	198	DPEIQVGLVQYGSPPVHWSLGRTEKEEVYRAAKNLSRREGRETKTAQAIWVACTEGF	257		
Db	195	GPQIQVGIQYGEDAVHFEHLNDYRSVKDVVEASHIEQRGGTETRTAFGIEFARSEAF	254		
Qy	258	SQSHGGRPEARLLVVTVDGSHDGEELPAALKACEAGRVTRYGIAVLGHYLRQRDPSS	317		
Db	255	QK--GGRKGAKVMIVITDGSHPDLEKVIQSEKDNVTRYAVAVLGYNNRRGINPET	312		
Qy	318	FLRIRITASPPDERFFENVTDALTDIVDALGRIFLGEGSHAENESSFGLMSQIGF	377		
Db	313	FLNEIKYIASPPDKHFFNVVDALDKDIVDALGRIFFLEGTN-KNETSFGLMSQIGF	371		
Qy	378	STHRLKDGILFGMVGADWGSVLWLEGGHRLFPFRMALEDFPPALQNHAAVLYGSVSS	437		
Db	372	SSHVVEDGILLGAVGYDMGAVLKEYSAGKVIPHRESYLKEFPEELKNHAAVLYGYVTS	431		
Qy	438	MLLGGRRLLFUSGAPRFRHRKGVIAFOLKDGAVVVAQSLQGEQIGSYFSGELCPLDTR	497		
Db	432	VWSSRQGRVYVAGAPRFNHTKVLFSMHNRSLLTIHQALRGEQIGSYFSGEITSVDVND	491		
Qy	498	DGTTDLVLLVAPMELGPNKGTGVYVLYVQOSSLTLTQGLQPEPP-QDARFGFMAL	556		
Db	492	DRVTDVLLVAGPMYFS--EGRGRKVVYVNL--RQNRVYVNGTLKDSHYSQNAFSGCSIASV	549		
Qy	557	PDLNQDGFADVAVGAPLEDHGQALYLYHGTSQSVRPHPAORIAAASMPHSAVYFGRSVD	616		
Db	550	QDLNQDSYNDVVVAGAPLEDHSRGAIYIPHGQTVILKKPMQRIITASELAPGLQHFQGSIH	609		

Qy	617	GRLLDGGDLVDVAVGAQGAAILSSRPVHLTPSLEVTPOAISVWVORDCRRRGQEAACL	676		
Db	610	QQLDLEDGLVDLAVGALGNVAVLWAPVQIINASHLFESKINIFHKDCRNGRDATCL	669		
Qy	677	TAALCFQVTSRTPEGRWDHQFYMRFTASLDEWTAGARAAFDGSGQRLSPRLRLSVGNVTC	736		
Db	670	AAFLCFPIPLAPHFQFATVGIYRNATMDERRYMPRAHLDEGGDQFTNRAVLSSGQEH	729		
Qy	737	EQHLHFVLDSDYLRPVALVTTFALDNTTTPGVPVNEGSPTSQKLPFSKDCGPDNECV	796		
Db	730	QRINFHVLDADYVVKPVAFVSVEYSLEDPDN--GPMLDNGWPTTLRLVSVPFWNGCNEDEHCV	788		
Qy	797	TDVLVQVMDIRSRK-----APFVVRGGRKVLVSTTLNRK	834		
Db	789	PDVLVDARSDLPFAMEYQCVLRPAQDCSSYLSFTTTFIESTRRRVAVEATLENRG	848		
Qy	835	ENAYNTSLSIFSRNLHLASLTPORESPIKVECAAPS--AHARLCSYGHVPVFTGAKVTF	892		
Db	849	ENAYSAVLNISQSENLFASLIQKDDSDNIECVNEBRRLLHKVCNVSYPPFRKAKAVAF	908		
Qy	893	LLFEFSCSLLSQVFGKLTASSDSLERNGLTQNTAQTSAITQYEPHLLFSSESTLHRY	952		
Db	909	RLDFEFSKSVFLHHLQIHLGAGSDSHEQDSTADNTALLRFLKYEADVLFTRSSLSHF	968		
Qy	953	EVHPYGTLPV--GPGPEFKTLRT-----IPEEEK	1125		
Db	969	EVKANSLESYDGIQPPFNCVQVQNLGFFPIHGVMMKIITVIATRGNRLLMLRDFTD	1028		
Qy	975	--NNASC--IVQNLTEPPGPPVHPBELOHTNRLNGSNTQCQVVRCHLQOLAKGTEVSVGLLR	1032		
Db	1029	QGNSTCNINWNSYEYRSTPTE--EDLSHAPQNHNSNDVSIICNL--RLASQETSFYLVG	1086		
Qy	1033	LVNNEFFRRAKFKSLTVVSTFELGTEGSLVQLTEASRWSESLLEVQTRPILSLMILI	1092		
Db	1087	NLWLTSLKALKYRSKLTIVNAALQRFSPFIFREEDPSRQVTFEISKQEDWQVPIWIV	1146		
Qy	1093	GSVLGGLLALLVFLCWLKLGFFFAHK-----IPEEEK	1125		
Db	1147	GSTLGGLLALLVLLWKLGGFFKSAKRREPGGLGPIPKELK	1188		
RESULT 4					
ID	Q7TQC3	PRELIMINARY;	PRT;	1188	AA.
AC	Q7TQC3;				
DT	01-OCT-2003	(TrEMBLrel. 25, Created)			
DT	01-OCT-2003	(TrEMBLrel. 25, Last sequence update)			
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)			
DE	All integrin.				
GN	Name=Itgall;				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	PubMed=1518724; DOI=10.1016/j.ydbio.2004.03.006;				
RA	Popova S.N., Rodriguez-Sanchez B., Liden A., Betscholtz C.,				
RA	van den Bos T., Gullberg D.,				
RT	"The mesenchymal alphal1beta1 integrin attenuates PDGF-BB-stimulated				
RL	chemotaxis of embryonic fibroblasts on collagens.";				
RN	Dev. Biol. 270:427-442 (2004).				
RP	SEQUENCE FROM N.A.				
RA	Johansson M., Popova S.N.;				
RL	Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.				
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).				
CC	-1- SIMILARITY: Belongs to the integrin alpha chain family.				
DR	EMBL; AY124460; AAM62130.1; -				
DR	HSP; P18614; 1CK4.				
DR	GO; GO:0006929; P:substrate-bound cell migration; IMP.				
DR	InterPro; IPR000413; Integrin_alpha.				
DR	InterPro; IPR002035; VWF_A.				



```
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00092; VWA; 1.
DR SMART; SM00191; Int alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
KW Cell adhesion; Integrin; Transmembrane.
SQ SEQUENCE 1188 AA; 133067 MW; 69B2BF20D75E4271 CRC64;

Query Match          39.3%; Score 2333.5; DB 2; Length 1188;
Best Local Similarity 42.2%; Pred. No. 4e-158;
Matches 499; Conservative 200; Mismatches 398; Indels 85; Gaps 15;

QY 18 GLCSPPNLDHHPRLPFGPPEAEFGYSLVQVGGQRMVLVGPMDGPGSDRRGDVYRCP 77
DB 18 GFTDTFNMDRNPRVIAGPSAAPPFGYTVQQHDISGKKLVVVGADMETNGHQTKGDVYKCP 77
QY 78 VGGHNAFCAKGLHDYQLGNSHPAUNMLHGLMSLLETGDGGGPMACAPLMSRACGSSVF 137
DB 78 VTOGN---CTKLNIGRVTLGNVSRKDNMRGLGSLATNPKNDSPLACSLWSHECGSSYY 134
QY 138 SSGICARVDASFOQGS LAPTAQRCPTMYMDVIVLDGNSIYPMSEVQTFLLRLRLVGKLF 197
DB 135 TTGCSRVNSFRSKTVAPALQRCQTYMDIVIVLDGNSIYPMSEVQTFLLRLRLVGKLF 194
QY 198 DPEQIQVLQVQSGSPVHWSLGDFTKKEEVRAAKNLSRRREGRTKTAQAIMVACTEGF 257
DB 195 GPGQIQVGIQVQGEDAHEPHLNDYRSKDVVEAASHIEQRGGTETRTAFGIEPARGSEAF 254
QY 258 SQSHGGPEARLLVVVTDGESHGDEELPAALKAACAGRVTRYGIAVLGHLRYLRQRPSS 317
DB 255 QK--GGRKGAKVMIVTDGESHDSPLDKVIROSEKDNTRYAVAVLGYNNRGINPET 312
QY 318 FLRIRTIASDPDRFPFNVTDEAALTDIVDALGDRIFGLEGSHAENESSFGLMSOIGF 377
DB 313 FLNEIKTIASDPDKHFNVTDEAALDKDIVDALGDRIFSLGNTN-KNETSFGLMSQTGP 371
QY 378 STHRLKDGILFGMVGAYDMGQSVLWLEGHRLFPFRMALEDEFPFALQNAAYLGYSVSS 437
DB 372 SSVHVEDGILLGAVGAYDMGAVLKETSAGKVI PHRESYLKEPPEELKNHVAIYLGTVTS 431
QY 438 MLLRGGRLFLSGAPRFRHKGKVIATFQKKGAVRVAQSLQGEQIGSYFGSELCPDTR 497
DB 432 WSSRQGRVYVAGAPRNHTGKVLFLSMHNNRSITIHQALRGEQIGSYFGSEITSVDVND 491
QY 498 DGTDDLVAAPMFLPQNKETGRVYVYVVGQQLTLTQGTLOPEPP-QDARFGFAMGAL 556
DB 492 DRVTDVLLVAGPMYFS-EGRGRKVYVNL-RQNFVYNGTLKDSHYSQNAFQSCIASV 549
QY 557 PDLNQDGFADVAVGAPLEDGHQALYLYHGQSGVRPHPAQRIAAASMPHALSYFGRSVD 616
DB 550 QDLNQDSYNDVVVGAFLDSDHRGAIYIFHGPTWILKKPMQORITASELAPGLQHFQCSIH 609
QY 617 GRLLDGDGLVDVAVGACGAAILLSRPVHLTSLPSLEVTPQASVVDORCRRQEAACL 676
DB 610 GQLDNLDEGLVDLAVGALGNAVLWAPVVPVQINASLHFEPSKINIFHKDCKNRGDATCL 669
QY 677 TAALCFQVTSRTPGRDHQFVMTFASLDEWTAGARAAFDGSGORLSPRRRLRSLGVNVC 736
DB 670 AAFLCFPIFLAFHQTATVGIIRYNATWDERYMPRAHLEGGQFTNRAVLSSGQEH 729
QY 737 EQLHFLHVDTSYLRLPVALVTFALDNTTKPGPVINEGSPTSIQKLVPFSDKDCPDNECV 796
DB 730 QRINFHVDLTADYVKVAFSVEYSLEDPDN-GPMLDNGWPTTLRVSPVFWNCGNDEHCV 788
QY 797 TDLVLQVMDIRGRK-----APVVGRRKVLVSTLTLENRK 834
DB 789 PDLVLDARSDLPTAMEYQCQVLRGPAQDCSSYTLSPDTTFFIESTRRRAVEATLENRG 848
QY 835 ENAYNTSLIIFSNLHLASTPQESPIKVECAAPS--AHARLCSVGHVPFQGAQVTF 892
DB 849 ENAYSAVLNTSQSENLOFASLIQKDDSDNSECYNEERLHKVCNVSYPPFRKAKAVAF 908
QY 893 LLEFEFSCSLLSQVFGKLTASSDSLENGTLQENTAQTSAYIQYEPHLLFSSSESTLHRY 952
```



DR PDB: 1MHP; X-ray; A/B=169-360.  
 DR RGD; 2923; Itgal.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 3.  
 DR Pfam; PF00357; Integrin\_alpha; 1.  
 DR Pfam; PF00092; VWA; 1.  
 DR PRINTS; PR01185; INTEGRINA.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SMO0191; Int\_alpha; 5.  
 DR SMART; SMO0327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS0234; VWA; 1.  
 DR 3D-structure; Calcium; Cell adhesion; Glycoprotein; Integrin;  
 KW Magnesium; Receptor; Repeat; Signal; Transmembrane.  
 FT CHAIN 1 28  
 FT DOMAIN 29 1180 Integrin alpha-1.  
 FT TRANSMEM 1143 1165 Extracellular (Potential).  
 FT DOMAIN 1166 1180 Potential.  
 FT REPEAT 104 103 Cytoplasmic (Potential).  
 FT REPEAT 175 388 FG-GAP 1.  
 FT REPEAT 377 432 VWA. FG-GAP 2.  
 FT REPEAT 433 484 FG-GAP 3.  
 FT REPEAT 485 565 FG-GAP 4.  
 FT REPEAT 567 626 FG-GAP 5.  
 FT REPEAT 629 681 FG-GAP 6.  
 FT CA\_BIND 497 505 Potential.  
 FT CA\_BIND 579 587 Potential.  
 FT CA\_BIND 641 649 Potential.  
 FT SITE 1168 1172 GFFKR motif.  
 FT DISULFID 82 92 By similarity.  
 FT DISULFID 687 696 By similarity.  
 FT DISULFID 702 755 By similarity.  
 FT DISULFID 807 813 By similarity.  
 FT DISULFID 877 885 By similarity.  
 FT DISULFID 1029 1062 By similarity.  
 FT DISULFID 1066 1073 By similarity.  
 FT CARBOHYD 100 100 N-linked (GlcNAc. . .)  
 FT CARBOHYD 105 105 N-linked (GlcNAc. . .)  
 FT CARBOHYD 112 112 N-linked (GlcNAc. . .)  
 FT CARBOHYD 217 217 N-linked (GlcNAc. . .)  
 FT CARBOHYD 317 317 N-linked (GlcNAc. . .)  
 FT CARBOHYD 341 341 N-linked (GlcNAc. . .)  
 FT CARBOHYD 402 402 N-linked (GlcNAc. . .)  
 FT CARBOHYD 418 418 N-linked (GlcNAc. . .)  
 FT CARBOHYD 459 459 N-linked (GlcNAc. . .)  
 FT CARBOHYD 531 531 N-linked (GlcNAc. . .)  
 FT CARBOHYD 698 698 N-linked (GlcNAc. . .)  
 FT CARBOHYD 747 747 N-linked (GlcNAc. . .)  
 FT CARBOHYD 779 779 N-linked (GlcNAc. . .)  
 FT CARBOHYD 820 820 N-linked (GlcNAc. . .)  
 FT CARBOHYD 839 839 N-linked (GlcNAc. . .)  
 FT CARBOHYD 882 882 N-linked (GlcNAc. . .)  
 FT CARBOHYD 907 907 N-linked (GlcNAc. . .)  
 FT CARBOHYD 938 938 N-linked (GlcNAc. . .)  
 FT CARBOHYD 965 965 N-linked (GlcNAc. . .)  
 FT CARBOHYD 973 973 N-linked (GlcNAc. . .)  
 FT CARBOHYD 1007 1007 N-linked (GlcNAc. . .)  
 FT CARBOHYD 1084 1084 N-linked (GlcNAc. . .)  
 FT CARBOHYD 1103 1103 N-linked (GlcNAc. . .)  
 FT CARBOHYD 1114 1114 N-linked (GlcNAc. . .)  
 FT STRAND 172 178  
 FT HELIX 186 197  
 FT TURN 198 199  
 FT TURN 204 205  
 FT STRAND 208 214  
 FT STRAND 218 222  
 FT TURN 224 226  
 FT TURN 230 238  
 FT TURN 239 239  
 FT HELIX 250 259

FT TURN 260 261  
 FT HELIX 264 266  
 FT TURN 267 267  
 FT TURN 270 271  
 FT STRAND 274 280  
 FT TURN 285 286  
 FT HELIX 287 289  
 FT HELIX 290 299  
 FT TURN 300 301  
 FT STRAND 302 309  
 FT HELIX 311 315  
 FT TURN 316 317  
 FT HELIX 321 328  
 FT TURN 329 330  
 FT HELIX 335 338  
 FT STRAND 339 342  
 FT HELIX 346 350  
 FT HELIX 352 359  
 SQ SEQUENCE 1180 AA; 130808 MW; 8E5DA2BE02362EE4 CRC64;  
 Query Match 31.8%; Score 1884; DB 1; Length 1180;  
 Best Local Similarity 36.2%; Pred. No. 7.6e-126;  
 Matches 432; Conservative 219; Mismatches 432; Indels 112; Gaps 25;  
 QY 13 LVFLTGLCSPLNLDHHPRLPPGPEAEFGYSVLQHVGGGQRMWLVGAPWDPSPGSDREGD 72  
 DB 19 LTVLGFVSNFVNDVRKMSFGSPVDMFGYTVQOYENEGKWLIGSLPLVGQPKARTGD 78  
 QY 73 VYRCPVGGAHNAPCAKAGHLGDYQLGNSSHPAV-----NMHLGMSLLETGDDGGFMACAPL 127  
 DB 79 VYKCPVGRERAMPCKLDLP-----VNTSIPNVTIKENMTFG-STLVTPNPGGFLACGPL 133  
 QY 128 WSRACGSSVFSSGICARVDASFPQGS LAPTAQRCPTVMVIVLDGNSNITYPWSVQTF 187  
 DB 134 YAYRCGHLHTTGICSDVSPTFQVNSFAP-VQECSTQLDIVIVLDGNSNITYPWSVIAF 192  
 QY 188 LRLVGLKFLDPEQIOVLVQYGESPVHWSLGFRTKEEVVRAAKNLSREGRETKTAQ 247  
 DB 193 LNDLLKRMKGPKQTVQVIGENVTNFKNSYSTEELVVAANKIGRGGIQTMTAL 252  
 QY 248 AIMVACTGFSQSHGGRPEARLLVVTGDSHDEBELPAALKACEAGRVTRYGIIVLGH 307  
 DB 253 GIDTARKEATEARGARGVKVMVIVTDSHDNYRLKQVIOCDENIQRFSAIULGH 312  
 QY 308 YLRQRDPSPFLREIRTIASDPDRFFNVTVDEAALTIDVDALGDRIFGLGSGHAENESS 367  
 DB 313 YNRGNLSTKRFVEBIKSIASEPTEKHFNVSDDELAVTIKALGERIFALEATADQSAAS 372  
 QY 368 FGLEMSQIGPSTHRLKDGILFGMVGAYDWGGSVLWLSGGHRLFPFRMALEDEFPFALQNH 427  
 DB 373 FEMEMSTGFSAHYSQDWMVLMGAVYDWNGVTVMQKANKQVIPHNTTTFQTE--PAKWN 430  
 QY 428 --AAYLVGSYSSMLLRGRRLFLSGAPFRHRGKVIAPOLKKDGAVRVAQSLQGEQIGSY 485  
 DB 431 PLASYLGYTVNSATIPGD-VLYIAGQPRYNHTGVVIYKM-EDGNINILQTLGGEQIGSY 488  
 QY 486 FGSELCLDTRDGTDTDLVLAAPMFLGPQNKETGRVYVYLVGQSLTLTLOGTLQP---- 541  
 DB 489 FGSVLTITIDKDSYTDLLLVGAPMVGTEKEQGVVYAV-NQTRFEYQMSLEPIRQT 547  
 QY 542 -----EPQDARFGFANGALPDNLQDGFADVAVGAPLEDHGGALYLYHG 586  
 DB 548 CCSSLKDNCSCTKENKNEPCGARFGTAIAVKDLNVDGNDVVI GAPLEDHAGAVIYHG 607  
 QY 587 TQSGVRPHPAQRIAAASMPHALSYFGRSVDGRDLDDGLVDVAVGAQAAILSSRPV 646  
 DB 608 SGKTIREAYAQRIPSGGDGKTLKFGOSIHGEMDLNGDGLTDVTIGLGGALFWARDVA 667  
 QY 647 HLTPSLEVTPOAISVQDRCRRRQGEAVCLTAALCFQVTSRTPGRWHDQFYMRFTASLDE 706  
 DB 668 VVKVTMNFEPKNVNIQKNCRVECKEIVCINATMCFHVKLKSKEDSIYEADLQYRVTLDS 727  
 QY 707 WTAGARAAFDCSGQORLSFRRLRLSVGNVTCEQLHFLHVLDTSDYLRPVALVTFFALDNTTK 766

Db 728 LQISRSFFSGTQERKIQR--NITVRESECIHRSFYMLDKHDFQDSVRVTLDLFD---NLTD 782  
Qy 767 P--GVLNLEGSPTSIOKLVPSKDCGPDNECVTDLVQVNDIRGSRKAPVVRGGRKV 824  
Db 783 PENGVLDDALPNSVHEHIFPAKDCGKERCISDLTNVST----TEKSLIVKSHQDKP 838  
Qy 825 LVSTTLNKRKENAYNTSLISIFSRNLHLASLTP-QRESPIKVECAAPSAHARLCSVGHVP 883  
Db 839 NVSLTVNKGDSAYNTRTVVQHSNPLIFSGIEEIQKDS-----CESNQNIICRVGYPF 891  
Qy 884 FQCAKQVTFLEFPFSCSSLLSQVFGKLTASSDLSRNGTLQENTQATSAIYQEPHLFP 943  
Db 892 LRAGETVTFKIIQFNTSHLSENAIHLSATSDEEPLSLNDNEVNSIPVKYEVGLQF 951  
Qy 944 SSBSTLHRYEHPVGTLP-----VG-----PGRPFKTL----- 972  
Db 952 YSSASEHHISVAANETIPEFINSTEDIGNEINVYTIKRGHPMPBELQLQSIPNLTAD 1011  
Qy 973 -----RTNNAACIVQNLTEPPG-----PPVHPBELQHTNRLNGSNTQCQVVR 1015  
Db 1012 GYPVLYPIGWSSSDNVNCRPSLEDPGINSKGKWTISKSEVLKRGTIQDCSSTCGVATI 1071  
Qy 1016 HLGQLAKG-TEVSVGLRLVHNEPFRRAKFKSLTVSTFELGTGEGSVLQLTASRWSES 1074  
Db 1072 TCSLLPSDLSQVNVSL--LLWKPTFIRAHFSSNLTLRGELKSENSS--LTLSNRRKREL 1128  
Qy 1075 LLEVQVQ-RTLLISLWILGSLVGLGLLLALLVCLWKLGFPAHKKIPSEKREE 1128  
Db 1129 AIQISKDGLPGRVPLWILLISAFAGLLLLMLLILALWKIGFF---KRPLKKKMEK 1180

RESULT 6  
ITAL\_HUMAN  
ID \_ITAL\_HUMAN STANDARD; PRT; 1151 AA.  
AC P56199;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 25-OCT-2004 (Rel. 45, Last annotation update)  
DE Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a).  
GN Name=ITGAL;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93155124; PubMed=8428973;  
RA Briesewitz R., Epstein M.R., Marcantonio E.E.;  
RT "Expression of native and truncated forms of the human integrin alpha  
1 subunit";  
RL J. Biol. Chem. 268:2989-2996(1993).  
CC -!- FUNCTION: Integrin alpha-1/beta-1 is a receptor for laminin and  
collagen. It recognizes the proline-hydroxylated sequence G-P-P-G-  
E-R in collagen.  
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-1  
associates with beta-1.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- DOMAIN: The integrin I-domain (insert) is a VMPA domain. Integrins  
with I-domains do not undergo protease cleavage.  
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.  
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.  
CC -!- SIMILARITY: Contains 1 VMPA domain.  
CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD49a entry;  
WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd49a.htm".  
CC PIR; A45226; A45226.  
DR PDB; 1QC5; X-ray; A/B=140-333.  
DR PDB; 1QCY; X-ray; A=141-333.  
DR Genew; HGNC:6134; ITGA1.  
DR MIM; 192968;  
DR GO; GO:0008305; C:integrin complex; TAS.  
DR GO; GO:0005518; P:collagen binding; TAS.  
DR GO; GO:0007160; P:cell-matrix adhesion; NAS.

DR InterPro; IPR000413; Integrin\_alpha.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF01839; FG-GAP 3.  
DR Pfam; PF00357; Integrin\_alpha; 1.  
DR Pfam; PF00092; VWA; 1.  
DR PRINTS; PRO1185; INTEGRINA.  
DR PRINTS; PRO0453; VWFADOMAIN.  
DR SMART; SM00191; Int\_alpha; 5.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
DR PROSITE; PS00244; VWF\_A; 1.  
KW 3D-structure; Calcium; Cell adhesion; Glycoprotein; Integrin;  
Mgnesium; Receptor; Repeat; Transmembrane.  
FT DOMAIN 1 1113 Extracellular (Potential).  
FT TRANSMEM 1114 1136 Potential.  
FT DOMAIN 1137 1151 Cytoplasmic (Potential).  
FT REPEAT 16 75 FG-GAP 1.  
FT REPEAT 76 75 FG-GAP 2.  
FT DOMAIN 147 360 VMPA.  
FT REPEAT 349 404 FG-GAP 3.  
FT REPEAT 405 457 FG-GAP 4.  
FT REPEAT 459 520 FG-GAP 5.  
FT REPEAT 540 599 FG-GAP 6.  
FT REPEAT 602 654 FG-GAP 7.  
FT CA BIND 470 478 Potential.  
FT CA BIND 552 560 Potential.  
FT CA BIND 614 622 Potential.  
FT SITE 1139 1142 GPFKR motif.  
FT DISULFID 54 64 By similarity.  
FT DISULFID 660 669 By similarity.  
FT DISULFID 675 728 By similarity.  
FT DISULFID 780 786 By similarity.  
FT DISULFID 850 858 By similarity.  
FT DISULFID 1002 1034 By similarity.  
FT DISULFID 1037 1044 By similarity.  
FT CARBOHYD 46 46 N-linked (GlcNAc...)  
FT CARBOHYD 72 72 N-linked (GlcNAc...)  
FT CARBOHYD 77 77 N-linked (GlcNAc...)  
FT CARBOHYD 84 84 N-linked (GlcNAc...)  
FT CARBOHYD 189 189 N-linked (GlcNAc...)  
FT CARBOHYD 289 289 N-linked (GlcNAc...)  
FT CARBOHYD 313 313 N-linked (GlcNAc...)  
FT CARBOHYD 374 374 N-linked (GlcNAc...)  
FT CARBOHYD 390 390 N-linked (GlcNAc...)  
FT CARBOHYD 432 432 N-linked (GlcNAc...)  
FT CARBOHYD 504 504 N-linked (GlcNAc...)  
FT CARBOHYD 671 671 N-linked (GlcNAc...)  
FT CARBOHYD 720 720 N-linked (GlcNAc...)  
FT CARBOHYD 752 752 N-linked (GlcNAc...)  
FT CARBOHYD 812 812 N-linked (GlcNAc...)  
FT CARBOHYD 855 855 N-linked (GlcNAc...)  
FT CARBOHYD 880 880 N-linked (GlcNAc...)  
FT CARBOHYD 887 887 N-linked (GlcNAc...)  
FT CARBOHYD 911 911 N-linked (GlcNAc...)  
FT CARBOHYD 938 938 N-linked (GlcNAc...)  
FT CARBOHYD 946 946 N-linked (GlcNAc...)  
FT CARBOHYD 980 980 N-linked (GlcNAc...)  
FT CARBOHYD 1045 1045 N-linked (GlcNAc...)  
FT CARBOHYD 1055 1055 N-linked (GlcNAc...)  
FT CARBOHYD 1074 1074 N-linked (GlcNAc...)  
FT CARBOHYD 1085 1085 N-linked (GlcNAc...)  
FT STRAND 143 150 N-linked (GlcNAc...)  
FT TURN 153 154  
FT TURN 158 158  
FT TURN 159 159  
FT HELIX 170 171  
FT TURN 174 174  
FT TURN 176 177  
FT TURN 179 186  
FT STRAND 190 194  
FT TURN 196 197  
FT TURN 202 211  
FT HELIX 222 231

FT	TURN	232	234	713	QERKQV--NITVRKSECTKHSFYMLDKHFDQDSVRITLDF---	767
FT	HELIX	236	238	777	TSIQKLVPSKDCGPDNECVTDVLQVNMDIRSGRKAPFVVRGRRKVLSTTLENKEN	836
FT	TURN	239	239	768	NSVHEYIPFAKDCGKNCISLISLHV---	823
FT	STRAND	242	243	837	AYNTSLSIIPSRNLHLASLTPQRESPIKVECAAPSAHARLCSVGHVPFQTCAGKVTFLLEF	896
FT	TURN	245	252	824	AYNTRIVHSPNLVFSGI---	877
FT	HELIX	257	258	897	EFCSLLSQVFGKLTASSDLSERNGTLOENTAGTSAYIQVPEHLLFSSESSTLHRYEHP	956
FT	TURN	261	271	878	QNTSYLMENVTVILSATSDEEPETLSNVVNISIPKVEVGLQFVSSASEVHISIAA	937
FT	HELIX	262	271	957	YGTLP-----VG-----	993
FT	TURN	272	273	938	NETVPEVINSTEDIGNEINIFYLIRKSGSPMPBLKLSIS-----	990
FT	STRAND	274	281	994	PEELQHTNRLN-----	1021
FT	HELIX	283	287	991	PTGLSSSENANCRPHIFEDPFSINSKGKMTTSTDHLKRGTLDCNCKPATITCNLTS-S	1049
FT	TURN	288	289	1022	KGTEVSVGLLRVHNEFFRRAKFKSLTVVSTFELGTSEGSVLQLTSEARWSESLEVVQT	1081
FT	HELIX	293	302	1050	DISQVNVSL--ILWKPTFIKSYFSSSLNLTIRGELRSENAS-LVLSSSNQKRELAIQISKD	1106
FT	HELIX	307	310	1082	RPILSLWLIGSVLGGILLALLVFLCWLKLPFAHKKIPEEBKREE	1128
FT	STRAND	311	314	1107	GLPGRVPLWVLLSAFAGLLLLMLLILALWKIGFF---	1151
FT	HELIX	317	323	RESULT 7		
FT	HELIX	324	330	O42094		
FT	SEQUENCE	1151	AA; 127837 MW; 6B3F3CIAABF52808 CRC64;	ID O42094		
Query Match				AC O42094		
Best Local Similarity				DT 01-JAN-1998 (TrEMBLrel. 05, Created)		
Matches				DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)		
426; Conservative 220; Mismatches 423; Indels 119; Gaps 23;				DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
Qy	23	FNLDEHPRLPPGPEAEFGYSVLQHVGGGQRMMLVGAPEWDPGSGDRRGDVYRCPVCGAH	82	DE	Alphal integrin.	
Db	1	FNVDKNSMTFSGVEDFMFGTVQOYENEEKVILGSLVPGQPKNRTGDIYKCPVGRGE	60	OS	Gallus gallus (Chicken).	
Qy	83	NAPCAKHGLGDYQIGNSHPAV-----NMHIGMSLLETGDDGPGMACAPLMSRACGSSVF	137	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Db	61	SLPCVKLDLP---VNTSIPNVTEVKENMTFG-STLVTPNPGGFLACGPLYAVRCGHLY	115	OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
Qy	138	SSGICARVDASFQPGSLAPTAQCPYMDVIVVILDGNSIYPHSEVQTFRLRLVGKLF	197	OC	Gallus.	
Db	116	TTGICSDVSPFTQVNVSIAP-VQECSTQLDVIIVLDGNSIYPNDSVTAFLNDLLKMDI	174	OX	NCBI_TaxID=9031;	
Qy	198	DPEIQIQLVQYSGSPVHWSLQDFRTKEEVVRAAKNLSRREGRETKTAQAIMVACTEGF	257	RN	(1)	
Db	175	GPQTQVIGVIGENVTHFNLNKSSTEEVLVAAKIVQGRQWTALGTDITARKEAF	234	RP	SEQUENCE FROM N.A.	
Qy	258	SQSHGRPEARLLVVVDGSHDGEELPAALKAACEAGRVTRYGIAGVLGHYLRQRDPSS	317	RC	TISSUE=Gizzard;	
Db	235	TEARGARRGVKKVMVITDGHSHDNRHLKKVIQDCEDENIQFSAIILGSYNRGNLSTEK	294	RX	MEDLINE=97476270; PubMed=9334246; DOI=10.1074/jbc.272.42.26643;	
Qy	318	FLRIRTIASPPDERFPFNVTDAAALTDIVDALGDRIFGLEGSHAENESSFGLMSQIGF	377	RA	Obata H., Hayaehi K., Nishida W., Momiyama T., Uchida A., Ochi T.,	
Db	295	FVEIKSIASEPTKEHFNVDDELALVTIVKTLGERIFALEATADQSAASFEMEMSQTF	354	RA	Sobue K.;	
Qy	378	STHRLKDGILFGMVGADWGSVLWLEGGHRLFPFPMALDEFPALQNHAAYLGYSVSS	437	RT	"Smooth muscle cell phenotype-dependent transcriptional regulation of	
Db	355	SAHYSQDWMLGAVGADWNGTVVWQKASQIIIPNTTFNVSTKKEPLASYLYGVNS	414	RT	the alphas integrin gene."	
Qy	438	MLLGRRLFLSGAPRFRHRGKIAFQKKDGAVRVAQSLQGEQIGSYFGSELCPDLTDR	497	RL	J. Biol. Chem. 272:26643-26651 (1997).	
Db	415	ATASSGDVLYIAGQPRYNHTQVLIYRM-EDGNIKILQTLSEQIGSYFGSLITLTDIDK	473	CC	-I- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).	
Qy	498	DGTTDLVLAAPMELPQNKETGRVYVVLVQSQSLLTLOGLQP-----	541	CC	-I- SIMILARITY: Belongs to the integrin alpha chain family.	
Db	474	DSNTDILLVAGPMYMGTEKEKQKVVYVAL-NQTRFEYQMSLEPIKQTCSSRQHSCTT	532	DR	EMBL; AB000470; BAA23160.1; -.	
Qy	542	---EPQDARFGFAMGALPDINODGFADVAVGAPLEDGHQCALVYHGTQSGVPHPAQR	598	DR	PIR; A55348; A55348.	
Db	533	ENKNEPCARGTAAVAVKDLNLDGNDIVIGAPLEDHGHGAVYIHGSGTKIRKFAQR	592	DR	GO; GO:0008305; C:integrin complex; IEA.	
Qy	599	IAAASMPHALSYGRSDRLDGLDVLVDVAVGAQGAAILSSRPVHLTPSLEVPQA	658	DR	GO; GO:0007160; P:cell-matrix adhesion; IEA.	
Db	593	IPSGDGLTKLAFQSGIHGENDLNGDGLTDVTIGLGAALFWSRDVAVKVTVNFEFNK	652	DR	GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.	
Qy	659	ISVVQDRCRRGQBAVCLTAALCFQVTSRTFGRWDHOFYMRFTASLDEWTAGAAAFDGS	718	DR	InterPro; IPR000413; Integrin_alpha.	
Db	653	VNIQKNCHMEGKETVCINATVCFEVLKSKEDTIYADLOYRVTLDSLRSISFSGT	712	DR	Pfam; PF01839; FG-GAP; 1.	
Qy	719	GQRLSPRLRLSVGNVCEQLHFFHVLDTSDVLRPVVALTVTALONTTKP--GPVLNKGSP	776	DR	Pfam; PF00092; VWFA; 1.	
				DR	PRINTS; PR01185; INTEGRINA.	
				DR	PRINTS; PR00453; VWFADOMAIN.	
				DR	SMART; SM00191; Int alpha; 4.	
				DR	SMART; SM00327; VWFA; 1.	
				DR	PROSITE; PS00242; INTEGRIN_ALPHA; 1.	
				DR	PROSITE; PS00234; VWFA; 1.	
				KW	Cell adhesion; Integrin; Transmembrane.	

[illegible]



RX MEDLINE=99308879; PubMed=2545729; DOI=10.1083/jcb.109.1.397;  
RA Takada Y., Hemler M.E.;  
RT "The primary structure of the VLA-2/collagen receptor alpha 2 subunit  
RT (platelet GPIa): homology to other integrins and the presence of a  
RT possible collagen-binding domain.";  
RL J. Cell Biol. 109:397-407(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Rieder M.J., Arnel T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,  
RA Rajkumar N., Toth E.J., Yi Q., Nickerson D.A.;  
RT "SealSnpS, NHLBI HUG6682 program for genomic applications, UW-  
RT FHCRC, Seattle, WA (URL: <http://pga.gs.washington.edu/>).";  
RL Submitted (MAY-02) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 168-368.  
RX MEDLINE=98019223; PubMed=9353312; DOI=10.1074/jbc.272.45.28512;  
RA Emsley J., King S.L., Bergelson J.M., Liddington R.C.;  
RT "Crystal structure of the I domain from integrin alpha2beta1.";  
RL J. Biol. Chem. 272:28512-28517(1997).  
RN [4]  
RP VARIANT GLU-534.  
RX MEDLINE=94043762; PubMed=7901236;  
RA Santoso S., Kalb R., Walka M., Kiefel V., Mueller-Eckhardt C.,  
RA Newman P.J.;  
RT "The human platelet alloantigens Br(a) and Brb are associated with a  
RT single amino acid polymorphism on glycoprotein Ia (integrin subunit  
RT alpha 2).";  
RL J. Clin. Invest. 92:2427-2432(1993).  
RN [5]  
RP VARIANT GLU-534.  
RX MEDLINE=20206009; PubMed=10744142;  
RA Kroll H., Gardemann A., Fechter A., Haberbosch W., Santoso S.;  
RT "The impact of the glycoprotein Ia collagen receptor subunit A1648G  
RT gene polymorphism on coronary artery disease and acute myocardial  
RT infarction.";  
RL Thromb. Haemost. 83:392-396(2000).  
CC -1- FUNCTION: Integrin alpha-2/beta-1 is a receptor for laminin,  
CC collagen, collagen C-propeptides, fibronectin and E-cadherin. It  
CC recognizes the proline-hydroxylated sequence G-P-G-E-R in  
CC collagen. It is responsible for adhesion of platelets and other  
CC cells to collagens, modulation of collagen and collagenase gene  
CC expression, force generation and organization of newly synthesized  
CC extracellular matrix.  
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-2  
CC associates with beta-1. Interacts with HPSS.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins  
CC with I-domains do not undergo protease cleavage.  
CC -1- POLYMORPHISM: Position 534 is associated with platelet-specific  
CC alloantigen HPA-5 (Br). HPA-5A/Br(a) has Lys-534 and HPA-5B/Br(b)  
CC has Glu-534. HPA-5B is involved in neonatal alloimmune  
CC thrombocytopenia (NAIT or NATP). The Lys-534-Glu polymorphism may  
CC play a role in coronary artery disease (CAD).  
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.  
CC -1- SIMILARITY: Contains 7 FG-GAP repeats.  
CC -1- SIMILARITY: Contains 1 VWFA domain.  
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD49b entry;  
CC WWW=<http://www.ncbi.nlm.nih.gov/prow/cd/cd49b.htm>.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; X17033; CAA34894.1; --  
CC EMBL; AF512556; AAM34795.1; --  
CC PIR; A33998; A33998  
CC PDB; 1A0X; X-ray; A/B=166-368.  
CC PDB; 1DZ1; X-ray; A=171-355.  
CC Genew; HGNC:6137; ITGA2.

DR MIM; 192974; --  
DR GO; GO:0008305; C:integrin complex; TAS.  
DR GO; GO:0005886; C:plasma membrane; TAS.  
DR GO; GO:0005518; P:collagen binding; TAS.  
DR GO; GO:0007596; P:blood coagulation; TAS.  
DR GO; GO:0007160; P:cell-matrix adhesion; TAS.  
DR GO; GO:0009887; P:organogenesis; TAS.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR InterPro; IPR002035; VWFA.  
DR Pfam; PF01839; FG-GAP\_2.  
DR Pfam; PF00357; Integrin\_alpha; 1.  
DR Pfam; PF00092; VWFA; 1.  
DR PRINTS; PR01185; INTEGRINA.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00191; Int\_alpha; 5.  
DR SMART; SM00327; VWFA\_1.  
DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
DR PROSITE; PSS0234; VWFA; 1.  
DR 3D-structure; Calcium; Cell adhesion; Direct protein sequencing;  
KW Glycoprotein; Integrin; Magnesium; Platelet; Polymorphism; Receptor;  
KW Repeat; Signal; Transmembrane.  
FT SIGNAL 1 29 Integrin alpha-2.  
FT CHAIN 30 1181 Extracellular (Potential).  
FT DOMAIN 30 1132 Potential.  
FT TRANSMEM 1133 1154 Cytoplasmic (Potential).  
FT DOMAIN 1155 1161 Interaction with HPSS.  
FT DOMAIN 1155 1161 FG-GAP 1.  
FT REPEAT 45 103 FG-GAP 2.  
FT REPEAT 104 378 VWFA.  
FT DOMAIN 188 378 FG-GAP 3.  
FT REPEAT 379 433 FG-GAP 4.  
FT REPEAT 434 486 FG-GAP 5.  
FT REPEAT 488 549 FG-GAP 6.  
FT REPEAT 551 610 FG-GAP 7.  
FT REPEAT 615 667 Potential.  
FT CA BIND 499 507 Potential.  
FT CA BIND 563 571 Potential.  
FT CA BIND 627 635 Potential.  
FT SITE 1157 1161 GPFKR motif.  
FT DISULFID 83 92 By similarity.  
FT DISULFID 680 737 By similarity.  
FT DISULFID 789 795 By similarity.  
FT DISULFID 865 876 By similarity.  
FT DISULFID 1019 1050 By similarity.  
FT DISULFID 1055 1060 By similarity.  
FT CARBOHYD 105 105 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 112 112 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 343 343 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 432 432 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 460 460 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 475 475 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 699 699 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 1057 1057 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 1074 1074 N-linked (GlcNAc... ) (Potential).  
FT CARBOHYD 1081 1081 N-linked (GlcNAc... ) (Potential).  
FT VARIANT 534 534 K -> E (in alloantigen HPA-5B;  
FT dbSNP:1801106).  
FT FTId=VAR\_003977.  
FT TURN 170 171  
FT STRAND 173 180  
FT TURN 183 184  
FT HELIX 188 199  
FT TURN 200 201  
FT STRAND 204 204  
FT TURN 206 207  
FT STRAND 209 216  
FT STRAND 220 224  
FT TURN 226 228  
FT HELIX 232 240  
FT TURN 241 241  
FT HELIX 252 262  
FT TURN 262 264  
FT HELIX 266 268

```
FT TURN 269
FT STRAND 275
FT HELIX 282
FT HELIX 291
FT HELIX 292
FT TURN 301
FT TURN 302
FT STRAND 303
FT STRAND 311
FT HELIX 313
FT TURN 317
FT TURN 318
FT TURN 319
FT HELIX 323
FT TURN 330
FT TURN 332
FT TURN 333
FT STRAND 340
FT STRAND 341
FT HELIX 344
FT HELIX 347
FT HELIX 353
FT TURN 354
FT TURN 362
FT TURN 363
SQ SEQUENCE 1181 AA; 129295 MW; 7E1B7ED968A94070 CRC64;

Query Match 29.28; Score 1733.5; DB 1; Length 1181;
Best Local Similarity 33.78; Pred. No. 4.9e-115;
Matches 410; Conservative 211; Mismatches 431; Indels 159; Gaps 32;

QY 11 LPLVFLTGL-----CSPFNLDHHPRLFPGLPPEAFGYSVLQHVGGGQRMVLVGPWD 63
DB 11 LPLLVLLSOGILNCCLAYNGLPEAKIFGSPSEQFVAVQQFIPNKGWLLVGPWS 70
QY 64 GPSGDRGVYRCPVGGAHNAPCAKHL-GDYQLGNSHPAVNMHLGMSLLETDGCGFM 122
DB 71 GFPENRMGVYKCPV-DLSTATCEKLNQTSIPNVTEMKTNMNLGLILTRNMGTCGFL 129
QY 123 ACAPLWSRACSSVFSGICARVDASFPQCSLAPTAQRCPYMDVIVLDGNSIYPWS 182
DB 130 TCGPLVAQQCGNYTTTGVGSDISDPDFQLSASFSPATQPCPSLIDVVVCDENSIYPWD 189
QY 183 EVOTFLRLVKLFIDPEQIQVLQVQGESPVHESWLSGLDFRTKEEVRAAKNLSRRREGRE 242
DB 190 AVKNLEKFKVQGLDGTGTQVGLIQYANPRVFNLTYYTKKEMIVATSQTSQYGGDL 249
QY 243 TKTAQAIWACTGESQSHGSRPEARLLVVVTVDGESHGDEELPAALKACEAGRVTRYGI 302
DB 250 TMTFGALQYARKAYSAAAGRRSATKVMVVVTDGESHGSMKXAVDQCNDHNLRFGI 309
QY 303 AVLGHYLRORDSSFLREITIASDPERFFNVNTEAALTDIVDALGRIFGLESHA 362
DB 310 AVLGYNRLNLDTKNLKEIKAIASIPTRYFFNVNDEAALKEKAGTGFISIEGT-V 368
QY 363 ENESSFGLMSQIGFST-HRLKDGILFGMVAGVWGCVSLW-LEGGHRLFPPRMALEDE 419
DB 369 QGGDNFQVMSQVGFSDYSSQNDILMLGAVGFCWGSTIVOKTSHGHLFP-----KQA 423
QY 420 FPPALQ--NHAAYLGVSVSMRLGRRRLFLSGAPRFRHKGKVIAPOLKKGAVRVQAQL 477
DB 424 FDQILQDRNHSSLYGVSAV-ISTGESTHFVAGAPRANYTQIVLYSVNENGNTIVIAH 482
QY 478 QGEQIGSYFSGELCPDTRDCTDVLVLAAPMFLGPNKQKTRGVYV-----LVGQSL 532
DB 483 RGDQIGSYFSGVLCSDVDKDTITDVLVAGAPMTMSDLKBEGRVYLFITKKGILGHQHF 542
QY 533 LTLQGTLPQEPDAPFGFAMGALPDNLNODGFADVAVGAPLEDHOGALYHGTQSGVR 592
DB 543 --LEG---PEGIENRFGSAALSDINMGDNVDIVGSPLENQNSGAVIYNGHQGTIR 597
QY 593 PHPAQIAAAA--SMPHALSIFGRSDGRDLDDGLVDVAVGAQGAAILSSRPVHLTP 650
DB 598 TKYSQKILGSDGAFRSHLYQFGRSLDGYGLNGDSITDVSIGAFGVQVQLMSQSIADVAI 657
QY 651 SLEVTPOAISVVQDRCRRRQGEANCLTALCFQVTSRTPGRWDHOFYMRFTASLDEWTAG 710
DB 658 EASPTPEKITLVNNAQ-----IILKLCFSAKFR-PTQNNQVAIVNITLDA----- 704
QY 711 ARAAFDGGGRLSPRL-----RLSGNV-----TCEQLHFHVLDTSDYLRVALTVT 758
DB 705 -----DGFSSRVTSGFLKENNERCLQKNMVMVNAQSCFEHIIXIQEPSDVVNSLDKVD 759
```

```
QY 759 FALDNTTKPG--PVLNEGSPSTISQKLVFSPKDCPDNECVTDLVQVNMDIRSKAPFV 816
DB 760 ISLEN---PGTSPALEYSETAKVFSPHFHKGDCGDLCTSDLVLDYR-QIPAAEQEOPFI 815
QY 817 VRGRRKVLVSTTLENKENAYNTSLSIIFSRNLHLASLTPQRESPI---KVEC-AAPSA 872
DB 816 VSNQKRLTFSVTLKNKRESAYNTGIVVDESENLFASFS-----LPVDGTEVTCQVAASQ 871
QY 873 HARLCSVGHVPVQTGAKVTFLEFEPFSCSLLSQVFGKLTASSDSLRNGTLOQNTAOTS 932
DB 872 KSVACDVGYPAKREQOVTFETINFDFNLQNLQNASLSFQALSSESQENKA--DNLVNLK 929
QY 933 AYIQYBPHLLFSSSESTLHRYEHPYCTLP-----VCPGPEKTTLTNNASCIVQ---- 982
DB 930 IPLLDAEHLTRSTNINFYEISSDGNVPSIVHSFEDVGPKFIFSLKVTTCGVSVPVMAV 989
QY 983 -----NLTEPPGP-----PVHP-----EELQHTNRLN 1004
DB 990 IHHPTVTEKKNPLMYLTGVQTDKAGDISCNADINPLKIGQTSSTSVFSKSENFRTKELN 1049
QY 1005 GSNTQCVQVRECHLQGLAKGTEVSVGLLRLVHNEFFRPAKFKSLTVVSTFELGTBEGSVLQ 1064
DB 1050 CRTASCNVTCWLDKDVHMKGEYFVNVTRIWNGTFASTFTQVQLTAAAEINTYNPEIV- 1108
QY 1065 LTEARSWESLLEVVQTRPILISLMI-----LIGSVLGLLLALLVFLCWLK 1111
DB 1109 -----VIEDNTVTIPLIMKPKDEKAEVPTGVIIGSIAGILLALLAILLAILWK 1155
QY 1112 LGFFAHK-----KIPEE 1123
DB 1156 LGFFKRYEKWTKNPDE 1172

RESULT 10
ITA2 MOUSE
ID ITA2 MOUSE STANDARD; PRT; 1178 AA.
AC Q62469; Q62163;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-2 precursor (platelet membrane glycoprotein Ia) (GPIa)
DE (Collagen receptor) (VLA-2 alpha chain) (CD49b).
GN Name=Itga2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX MEDLINE=94363406; PubMed=8081889;
RA Edelman J.M., Chan B.M., Uniyal S., Onodera H., Wang D.Z.,
RA Damjanovich L., Latzer D.B., Finberg R.W., Beigelson J.M.;
RT "The mouse VLA-2 homologue supports collagen and laminin adhesion but
RT not virus binding.";
RL Cell Adhes. Commun. 2:131-143 (1994).
RN [2]
RP SEQUENCE OF 450-1178 FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=94355691; PubMed=7521231;
RA Wu J.E., Santoro S.A.;
RT "Complex patterns of expression suggest extensive roles for the alpha
RT 2 beta 1 integrin in murine development.";
RL Dev. Dyn. 199:292-314 (1994).
CC -!- FUNCTION: Integrin alpha-2/beta-1 is a collagen receptor, being
CC responsible for adhesion of platelets and other cells to
CC collagens, modulation of collagen and collagenase gene expression,
CC force generation and organization of newly synthesized
CC extracellular matrix. It is also a receptor for laminins, collagen
CC C-propeptides and B-cadherin. Mice homozygous for a null mutation
CC in the alpha-2 die very early in embryogenesis.
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-2
CC associates with beta-1.
```



CC -|- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -|- DOMAIN: The integrin I-domain (insert) is a VWF domain. Integrins  
 CC with I-domains do not undergo protease cleavage.  
 CC -|- SIMILARITY: Belongs to the integrin alpha chain family.  
 CC -|- SIMILARITY: Contains 7 FG-GAP repeats.  
 CC -|- SIMILARITY: Contains 1 VWF domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z29987; CA82877.1; -;  
 DR EMBL; X75427; CA53178.1; -;  
 DR PIR; S44142; S44142.  
 DR HSP; P17301; IAOX.  
 DR MGD; MGI:96600; Itga2.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 2.  
 DR Pfam; PF00357; Integrin\_alpha; 1.  
 DR Pfam; PF00092; VWA; 1.  
 DR PRINTS; PR01185; INTEGRINA.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 5.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS0234; VWF; 1.  
 KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Platelet;  
 KW Receptor; Repeat; Signal; Transmembrane.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1178  
 FT DOMAIN 27 1129 Integrin\_alpha-2.  
 FT TRANSMEM 1130 1151 Extracellular (Potential).  
 FT DOMAIN 1152 1178 Potential.  
 FT REPEAT 42 100 Cytoplasmic (Potential).  
 FT REPEAT 101 2 FG-GAP 1.  
 FT REPEAT 185 2 FG-GAP 2.  
 FT REPEAT 375 2 VWF.  
 FT REPEAT 431 4 FG-GAP 3.  
 FT REPEAT 485 5 FG-GAP 4.  
 FT REPEAT 548 6 FG-GAP 5.  
 FT REPEAT 612 6 FG-GAP 6.  
 FT REPEAT 664 7 FG-GAP 7.  
 FT CA\_BIND 496 504 Potential.  
 FT CA\_BIND 560 568 Potential.  
 FT SITE 624 632 Potential.  
 FT SITE 480 482 Cell attachment site (Potential).  
 FT SITE 1154 1158 GPPKR motif.  
 FT DISULFID 80 89 By similarity.  
 FT DISULFID 677 734 By similarity.  
 FT DISULFID 786 792 By similarity.  
 FT DISULFID 862 873 By similarity.  
 FT DISULFID 1016 1047 By similarity.  
 FT DISULFID 1052 1057 By similarity.  
 FT CARBOHYD 102 102 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 109 109 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 429 429 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 457 457 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 472 472 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 696 696 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1054 1054 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1071 1071 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1078 1078 N-linked (GlcNAc... ) (Potential).  
 SQ SEQUENCE 1178 AA; 128926 MW; 1F194B9C0240F465 CRC64;  
 Query Match 28.9%; Score 1712.5; DB 1; Length 1178;  
 Best Local Similarity 34.7%; Pred. No. 1.6e-113;  
 Matches 412; Conservative 217; Mismatches 457; Indels 103; Gaps 32;  
 QY 9 LFPLVLFLT-GL---CSPFNLDHHPRLFPGPPEAFGYSVLQHVGGGQRMWLVGAPWDGP 65

Db 10 LLQLLMLVQGIINCLAYNVGLFGAKIFSGPSSEQFGYSVQQLTNQGNWLLVGSWSGF 69  
 QY 66 SGDRRGDVVRCVPVGGAAHNAKCAKGH- GDYQLGNSSHPAVNMHLGMSLLELTDGCGGPMAC 124  
 Db 70 PENRMGDVYKCPV-DLPTATCEKLNQNSASISNVTIKTNMSLGLTLTRNPTGGGLTC 128  
 QY 125 APIWSRACSSSVSSGICARVDASFPQGS LAPTAQRCPYMDVIVLVDGNSIYPMSEV 184  
 Db 129 GPLMAHQCGNYATGICSDVSPDFQLASFSPAVQACSLVDVVVCDNSIYPEAV 188  
 QY 185 QTFRLRLVGLKFTDPQIQVGLVQGESPVHWSLGDFTKSEVRAAKNLSRREGRETK 244  
 Db 189 KNFLVFPVTLGLDGPKKTKQVALIQYANEPRIIFNLNDFETKEDVMQVATSETROGGDLTN 248  
 QY 245 TAOAIMVACTEGFSQSHSGRPEARLLVVVTCDESHDGEELPAALKACAGRVTRYGIIV 304  
 Db 249 TFRALFARDYAYQTSGGRPATKVMVVVTDGESHGSKLTVIOQNDDELIRFGIIV 308  
 QY 305 LGHYLRQRDPSSFLREIRTIASDPDRFFNFVNTDEAALTDIVDALQDRIFLGGSHAEN 364  
 Db 309 LGYLNALDITKNLKEIKAIASPTPTERYFFNVADAEALKEKAGTLGEQIFSIETG-VQG 367  
 QY 365 ESSFGLEMSQIGPSTHRL--KQGLFGMVGAYDWGSGVLMLEGHR--LPFPRMALEDEF 420  
 Db 368 GDNFQEMAQVGFSAADYAPQNDILMLGAVGAFQWSGLV-QETSHKVPVIPP---KQAF 421  
 QY 421 PPALO--NHAAYLGYSVSSMLLGGRRFLSGAPRFRHGRKVIAPQLKDGAVRAVQSLO 478  
 Db 422 QVQLQDRNHSSFLGYSVAALSTEDGVH-FVAGAPRANYTQGVLYSVNKGNTVVIQSHR 480  
 QY 479 GEQIGSYFGSELCPLOTDRTDRTDVLVLAAPMFLGPQNKETGRVYVYLVQOQSLLTLQGT 538  
 Db 481 GDQIGSYFGSVLCSVDVDKDTITDVLVGPATYMDLKEEGKVLYLTITKGLNHQFPL 540  
 QY 539 LQPEPQDARFGPAMGALPDINODGPADVAAGAPLEDHOGALYLYHGTQSGVRPHPAQR 598  
 Db 541 EGPEGTGNARFGSAIAALSINMDGFNDVTVGSPVENENSGAVIYVNGHQGTIRTKYSQK 600  
 QY 599 IAAA--SMPHALSYFGSRVDGRLLDGDGLVDVAVAGQAAILLSRPIVHLTPSLVTP 656  
 Db 601 ILGNGAFRRHLQFFGRSLDGYGLNGDSITDVSIGALQVQLMSQSIADVAIEALFTP 660  
 QY 657 QAISVVQDRCRRRQGEAVCLTAALCFQVTSRTSRGRMDHQFYMFTASLDLDTWAGAAAFD 716  
 Db 661 DKITLTKDAK-----ITLKLCPRAEPRPAGQ-NNVAAILFNMTLDADGHSSRVTSR 711  
 QY 717 GSGQRLSPRLR--LSVGNV-TCEQLHFHVLDTSDYLRPVALVTVPALDNTKPG--PVL 771  
 Db 712 GVFPRENERPLQNMVNVQKSEHHISIQKPSDVVNPLDLRDVDSLEN---PGTSPAL 768  
 QY 772 NEGSPTSIQKLVPSKDCGPDNECVTDVLQVNMDIRGSKAPVVRGRRKVLVSTTLE 831  
 Db 769 EAYSETVKVFSIPFYKECGSDGICISDLIDV-QQLPAIQTSFIVSNQNKRTFFSVILK 827  
 QY 832 NRKENAYNTSLSIIFSRNLHLASITPQRESPI---KVECAAPSAHARL-CSVGHPVPTG 887  
 Db 828 NRGESAVTVVLAEPSENLFASFS---MPVDGTEVTCVGSQSKSVTCVGVPAKSE 883  
 QY 888 AKVTFLLEFPFGSSLLSVFGKLTASSDSLENGTLQENTAOISAVIQVEPHLLPSSS 947  
 Db 884 QQVTFITNFDNLQNLQNAAINFOAFSESQETNKA--DNSVSLTIPLLDAELHLTRST 941  
 QY 948 TLHRYEVHPYGTLP-----VGRG-----PEEK----- 969  
 Db 942 NINPYESSENAPSVIKSVEDIGKPIFSLKVTAGSAPVSMALVTTHIQYTKKNPLL 1001  
 QY 970 --TTLRTNNA---SCI--VQNLTEP---PGPPVHPPELQHTNRLNGSNTQCVVRCCHLQ 1019  
 Db 1002 YLTGIQTDQAGDISCTAEINPLKLPHTAPSVSPKFNENFRHTKELDCRTTSCNITCWLD 1061  
 QY 1020 LAKGTEVSVGLLKVHNEFFRRAKPKSLTVSTVFELGTGEGSVLQLTEARWSLESLEV 1079



Db 1062 LHMKAEBFVNTVTRVNRTPAASFTQVOLTAAAEIDTHNPQ-LFVIEENAVTIPLMIMK 1120  
 Qy 1080 QTPILISLWILGSLVGLLLALLVFCWLKLGFF--AHKXI---PEE 1123  
 Db 1121 PTEKAEVPTGVIIIGSIAGIIALLAMTAGLWKLKLGFFKRYKKMGQNPDE 1169  
 RESULT 11  
 Q6P1C7 PRELIMINARY; PRT; 1178 AA.  
 AC Q6P1C7  
 DT 05-JUL-2004 (T-EMBLrel. 27, Created)  
 DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)  
 DE Integrin alpha 2.  
 GN NamesMusga2;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Eye;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Moore T., Max S.I., Wang J., Haieh F.,  
 RA Hopkins R.F., Jordan H., Moore K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Diatchenko L., Marusina K., Bontadi M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Pahay J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bonifard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skaleka U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Eye;  
 RA Straube R.;  
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -1- SIMILARITY: Belongs to the integrin alpha chain family.  
 DR EMBL; BC065139; AAH65139.1; -.  
 DR HSSP; P17301; 1A0X.  
 DR GO; GO:0005615; C:extracellular space; TAS.  
 DR GO; GO:0016021; C:integral to membrane; TAS.  
 DR InterPro; IPR00413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 2.  
 DR Pfam; PF00357; Integrin\_alpha; 1.  
 DR Pfam; PF00092; VWA; 1.  
 DR PRINTS; PR01185; INTEGRIN.  
 DR SMART; PR00453; VWFADOMAIN.  
 DR SMART; SM00191; Int alpha; 5.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS02234; VWFA; 1.  
 KW Cell adhesion; Integrin; Transmembrane.  
 SQ SEQUENCE 1178 AA; 128954 MW; 62FAEA820242A9B6 CRC64;

Query Match 28.9%; Score 1712; DB 2; Length 1178;  
 Best Local Similarity 34.58; Pred. No. 1.7e-113;  
 Matches 409; Conservative 218; Mismatches 455; Indels 102; Gaps 31;

Qy 13 LVPLTGL--CSPENLDEHHPRLPFGPEAEFGYSVLQGVGGORWMLVGAPWDGSPGDRR 70  
 Db 15 LMLVQGLNCLATNVLGPGAKIFSGSSSEFGISVQQLTNPQGNLWLGSPGSPENRM 74  
 Qy 71 GDVYRCPVGGAHNAPCAKAGHL-GDYQLGNSSHPAVNMHLGMSLLETDDGSGFACAPLWS 129  
 Db 75 GDVYKCPV-DLPTATCEKLNQNSASISNVTEIKTNMGLGLTRNPGTGSGFLTCGPLWA 133  
 Qy 130 RACGSSVFSGGICARVDASFPQGSGLAPTAQRCPTYMDVIVLDGNSIYPWSEVQTFELR 189  
 Db 134 HQCGNQYATGICSDVSPDFQFLTSPFPAQCPSLVDDVVVVCDESNIYPWEAVKQPLV 193  
 Qy 190 RLVGKLFIDPEQIQVGLVQGESPVHWSLGDPRTEEVVRAAKNLRRREGRETAKAI 249  
 Db 194 KFTVGLDIGPKTQVALIQIANEPRIIFNLNDFETKEDMVQATSETKHQGDGLTNTFRAI 253  
 Qy 250 MVACTEGFSQSHGSRPEARLLVVVTDGESHGDEELPAALKACEAGRVTRYGIYAVLGHYL 309  
 Db 254 EFARDYAYSQTSGRPGATKVMVMTDGDGSDGSKLKTIVQCNDDDEILRFGIYAVLYLN 313  
 Qy 310 RRORDSSFLREIRTASDPDERFFNFVTDEAALTDIVDALGDRIFGLEGSHAENESFG 369  
 Db 314 RNALDTKNLIKEIKATASTPTERYFFNVADEAALLEKAGTLGEQIFSIETG-VQGGNFQ 372  
 Qy 370 LEMSOIGFSTHRL--KDGILFGVMGVADWCGSVLWLEGGHR--LFPFPMALDEDEFPALQ 425  
 Db 373 MENAQVGFSAADYAPQNDILMLGAVGAFDWSGTLV-QETSHKPVIFP-----KQAFDQVLQ 426  
 Qy 426 --NHAAYLYGYSVSMILRGRRRLFLSGAPFRHRGKVIAPQLKDGAVRVAQSQQGQIG 483  
 Db 427 DRNHSSFLGYSVAAISTEDGVH-FVAGAPRANYTGQIVLYSVNKGQNVTVIQSHRGDQIG 485  
 Qy 484 SYFGSELCPIDTRDGTDTLLVAAMPFLGPQNKETGRVYLVGQOSLTLTQOTLOPEP 543  
 Db 486 SYFGSVLCSVDVDKDTITDVLVCGAPTYMNDLKEGKGYLFTITKGLNHQHQLEGPEG 545  
 Qy 544 PQARFGFAGALPDNLQDGFADVAVGAPLEDHGQALYLYHGFGOSVRHPAPQRIAAA- 602  
 Db 546 TGNARFGSAALASDINMGDFNDVGVSPVENENSGAVIYNGHQTIRTKYSOKILGNSN 605  
 Qy 603 -SNPHALSYFGRSVDGRDLDDGLVDVAVGAQAAILLSSRPVHVTTPSLEVTPQAISSV 661  
 Db 606 GAFERHLQFFGSLDGYGLNGDSITDVSIGALGOVQLMSQSIADVAIEALFPDKITL 665  
 Qy 662 VQDRCRRRGOEAVCLTAALCFQVTSRTPRGRWDHQFYMRFTASLDDEWTAGARAAFDGSGQR 721  
 Db 666 LNKDAK-----ITLCLCFRAEFEPAGQ-NNQVAILFNMTLDADGSHSRVTSRGVFRE 716  
 Qy 722 LSPRLR--LSVGNV-TCEQLHFHVLDTSDYLRPVALTVTFDLNTTKPG--PVLNKGSP 776  
 Db 717 NSERFLQKNMVMNEVQKCEHHISIQKPSDVANPLDLRVDISLEN---PGTSPALEAYSE 773  
 Qy 777 TSIQKLVPFSKDCGPDNECVTDLVQVNMIDIRSKAPFVVVGRGRRKVLSTLLENKEN 836  
 Db 774 TVKVFSIPFYKEGSDGICISDILLDV-QQLPAIQOSFIVSNQNKLTSTVLNKGES 832  
 Qy 837 AYNTLSLSIIFSRNLHLASLTPQRESPI---KVECAAPSAAHARL-CSVGHVPVFGAKVTF 892  
 Db 833 AYNTVLALAEFSENLFPASPS---MPVDGTEVTCESVGSQKSVTCVGVGPAKSEQQVTF 888  
 Qy 893 LLEFEPSCSLLSQVFGKLTASDSLSLRNGTLQENTAQTSAYIQYEPHLLPSSSESTLHRY 952  
 Db 889 TINFDFNLQNLQAAINFQAFSESQETNKA--DNSVSLTIPLLYDAELHLTRSTNIFY 946  
 Qy 953 EVHPYGTLP-----VGPQ-----PEFK-----TTL 972  
 Db 947 EISSDENAPSVIKSVEDIGPKFIFSLKVTAGSAPVSMALVTIHPQTKENKPNLLYLTGI 1006  
 Qy 973 RTNNA---SCI---VQNLTEP---PGPPVHPPEELQHTNRLNGSNQTCQVVRCHLQGLAGT 1024  
 Db 1007 QTDQAGDISCTAEINPLKPLHTAPSVSFKNENFRHTKELDCRTTSCSNITCWLKDLHKA 1066  
 Qy 1025 EVSVGLLRLVHNEFFRRRAKFKSLTVVSTFELGEEGSLVQLTEASRWSESLLEVVQTRPI 1084

```
Db 1067 EYFNVTRVNRNFAASTFTQVLTAARAAIDHTNPQ-LFVIBENAVTIFLMIMKPTKA 1125
QY 1085 LISLWILGGLGLLLALLVCLWKLGF--AHKKI---PBE 1123
Db 1126 EVPTGVIIIGIIAGILLALLTAGLWKLGFKKQYKKGQNPDE 1169

RESULT 12
Q8WY18
ID Q8WY18 PRELIMINARY; PRT; 823 AA.
AC Q8WY18; 2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE MSTP018.
GN Name=MST018;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Aorta;
RA Xu Y.Y., Sun L.Z., Wu Q.Y., Liu Y.Q., Liu B., Zhao B., Wang X.Y.,
RA Song L., Ye J., Sheng H., Gao Y., Zhang C.L., Zhang J., Wei Y.J.,
RA Sun Y.H., Jiang Y.X., Zhao X.W., Liu S., Liu L.S., Ding J.F.,
RA Gao R.L., Qiang B.O., Yuan J.G., Liew C.C., Zhao M.S., Hui R.T.;
RA Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the Integrin alpha chain family.
DR EMBL; AF111799; AAL39001.1; .
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0005155; F:protein binding; IEA.
DR GO; GO:0007160; F:cell-matrix adhesion; IEA.
DR GO; GO:0007229; F:integrin-mediated signaling pathway; IEA.
DR Pfam; PF01839; FG-GAP; 1.
DR PRINTS; PRO1185; INTEGRINA.
DR SMART; SM00191; Int alpha; 4.
KM Cell adhesion; Integrin; Transmembrane.
SQ SEQUENCE 823 AA; 92672 MW; D84E78079DCD4925 CRC64;

Query Match 23.7%; Score 1403; DB 2; Length 823;
Best Local Similarity 38.3%; Pred. No. 1.5e-91;
Matches 316; Conservative 142; Mismatches 293; Indels 74; Gaps 13;

QY 372 MSQIGFTHRLKQILFGMVGYDWGGSVLWLEGGHRLFPFRMALEDEFPALQNHAYL 431
Db 1 MSQTGFSSHVVEDGVLGAGVYDWNGAVLKVETSAKVIPLRESYLKFEPEELKXHGAYL 60
QY 432 GYSVSMLLRGGRFLPLSGAPRFRHRKVIATFQLKDGAVRVAQSLOGEQISYFGSEL 491
Db 61 GYTVTSVSSRQGRVYVAGAPRFNHTKVLFTMHNNRSLTIHQMGGQISYFGSEIT 120
QY 492 PLDTDRDGTDLVLAAPAFMELPGQNKETGRVYVYLVQGSLLITLQGLQPEPP-QDARFG 550
Db 121 SVDLDGSDVTLVVGAPMYFN-EGREGRKVVYVEL-RQNFVYNGTLKOSHSYQNAF 178
QY 551 FAMGALPDLNQDGFADVAVGAPLEDGCGALYVHGTSQGVRRPHPAQRIAAASMPHALSY 610
Db 179 SSIAVSRLDNQDSYNDVVVVGAPLEDNHAGAIYIFHGFSGILKTPKQIRITASELATGLQY 238
QY 611 FGRSDGRLDLDGDDLDVAVGAGAAILLSSRPVHLTPSLEVTPOAISVVQDCHRRG 670
Db 239 FGCISHQGLDLDGLDLDVAVGAGAAILLSSRPVHLTPSLEVTPOAISVVQDCHRRG 670
QY 671 QEAVCLTAALCFQVTSRTPGRDWDQFYMRFTASIDETAGARAAFDGSGQRLSPRLRLS 730
Db 299 RDATCLAAFLCFTPIFLAPHPQTITVGRYNATWDERRYTPRAHLDEGGDRFTNRAVLS 358
QY 731 VGNVTCQLHPLVLDTSYLPALVATVTFPDLNTPKGPVLNBSGSPSTSIQKLVPFSDKCG 790
Db 731 VGNVTCQLHPLVLDTSYLPALVATVTFPDLNTPKGPVLNBSGSPSTSIQKLVPFSDKCG 790

Db 359 SQQLCERINFHVLDTADYVYKVPVTFVSVEYSLEDPDH-GPMLDDGMPTTLRVSPVFMNGCN 417
QY 791 PDNECVTDLVQVNMID-----RGSRK-----APFVVRGGRKKVLVST 828
Db 418 EDHCVDPDLVDARSDLPTAMEYQVRVLRKPAQDCSAYTILSPDTTTFIESTQRVAVEA 477
QY 829 TLENRKENAYNTSLSIIFSRLHLASLTLPQRESPIKVECAAPS--AHARLCSVGHVPVOT 886
Db 478 TLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSIKVECNERRLQKQVCNVSYPPFRA 537
QY 887 GAKVTFLLEFPSCSLLSQVFKLTASSDSLNRNGTLQNTAQTSAIYQYEPHLLPSSSE 946
Db 538 KAKVAFRLDFEFSKSLFLHLELELAAGSDSNRSTKEDNVAPLPHLKYEADVLFTFS 597
QY 947 STLHRYEVHPYGTLP-VGPGPEFKTTLRTN----- 975
Db 598 SSLSHVEVKLNSLSERYDGGPPFCIFRLQNLGLPPIHGMMKIITPIATRSNRLKL 657
QY 976 -----NASC-IVQNLTPEPPGVHPPELQHTNRLNGSNTQCQVVVRCHLGLQAKGTEV 1026
Db 658 RDLTDEANTSCNIGWNSTEYRTPVE-EDLRRAPQLNHSNDVSVINCNI-RLVPPNQEI 715
QY 1027 SVGLLRLVHNEPFRRAKFKSLTVVSTFELCTEGSVLQLTASRNWSESLLEVVTQTRILI 1086
Db 716 NFHLGLNMLRSLSLKALKYKSMKIMVNAALQRFHSPFIFREEDPSRQIVFEISKQEDMQV 775
QY 1087 SLWLIGSVLGLLLALLVCLWKLGFPAHKKIPBEEKREKLE 1131
Db 776 PIWIVGSTLGGULLALLVCLWKLGFPSAR-----RRREPGLD 816

RESULT 13
Q8CE84
ID Q8CE84 PRELIMINARY; PRT; 823 AA.
AC Q8CE84; 2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
DE library, clone:4732459H24 product:MSTP018 homolog.
GN Name=Itgall;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RC MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RX Carninci P., Hayashizaki Y.;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RX RIKEN FANTOM Consortium;
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RC The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
```

```

RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RL prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin; DOI=10.1101/gr.152600;
RC MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuami T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hatanaka M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuura T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai K., Sakai K., Sakazume N., Sano H.,
RA Saeki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
DR EMBL; AK028821; BAC26137.1; --
DR MBL; MGI:2442114; Tcgal1.
DR GO; GO:0006929; P:substrate-bound cell migration; IMP.
DR InterPro; IPR00413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 1.
DR PRINTS; PR01185; INTEGRINA.
DR SMART; SM00191; Int_alpha; 4.
KW Cell adhesion; Integrin; Transmembrane.
SQ SEQUENCE 823 AA; 92264 MW; A330236324A0E089 CRC64;

Query Match 23.4%; Score 1388.5; DB 2; Length 823;
Best Local Similarity 38.0%; Pred. NO. 1.7e-90;
Matches 315; Conservative 138; Mismatches 296; Indels 79; Gaps 12;

QY 372 MSQIGFSTHRLKDGILFGMVCAYDGGSVLWEGHRLFPFRMALEDFFPALQNHAYL 431
DB 1 MSQIGFSSHVVEDGILLGAVGAYDNGAVLKVETSGAGKVPHPRESYLYKEFFPEELKNHAYL 60
QY 432 GYSVSMLLRGRLFLSLGAPFRHRGKVIAPQLKDGAVRVAOSLQEQIGSVFGSEL 491
DB 61 GYTVSVSSVSSQGRVYVAGAPFRHNTGKVLFLSMHNRSLTIHQALRGEIGSVFGSEIT 120
QY 492 PLDTDRGDTTDLVLAAPFMFLGPQNKETGRVYVYLVGQSSLLTLTQGLTQPEPP-QDARFG 550
DB 121 SVDVNDVRVTDVLLVAGPMYFS-EGRERGKVVYVNL-RQNRVYVNTGLKDSHSYQNARFG 178
QY 551 FMGALPDLNODGADVAVGAPLEDGOGALYLYHGTSQGVVRPAPQIAAASHPALSY 610
DB 179 SCIASVQDLNDSYNDVVVVGAPLEDGSHRGATYIFHGFTNTLTKRPMQRITASELAPGLQH 238
QY 611 FGRSVDGRLLDGDLDVAVGACAAITLLSSRPVHLTPSLEVTTPQASIVVQDCRRRG 670
DB 239 FGCSTHGQLDNEDGLVDLAVGALGNVVLWARPVQVNASLHFEPSKINIFHKDCRNG 298
QY 671 QEAVCLTAALCFQVTSRTPGRWDHQFYVMRFTASLDWTAGARAAFDGSGQLSPRLRLS 730
DB 299 RDATCLAAFLCFIPFLAPHPQTATVGRVYATWDERRYMPRAHLDEGDDQFTNRAVLIS 358
QY 731 VGNVTCQLHPLHLDTSYLRPVALVTFTALDNTTKPGFVLNKGSPSTSIQKLVPFSKDCG 790

```

```

Db 359 SGHEHCORINFHLDTRADYVKPAPSVSEYSLSDPDN-CPMLDNGWPTLTVSVFVWNGCN 417
QY 791 PDNECVTDLVQVNMDIRSRK-----APFVVRGGRKKVLYST 828
DB 418 EDEHCVPDLVDARSDDLPTAMEYQCVLGRPAQDCSSYTLSDFTTVFIESTRRRVAVEA 477
QY 829 TLENKRNANTSLSIIFSRNLHLASLTPORESPIKVECAAPS--AHARLCSVGHVPQOT 886
DB 478 TLENRGENAYSAVLINQSLENLOFASLIQDDSDNSIECVNEERLHKVCNVSYPFFRA 537
QY 887 GAKVTFLLEPEFSCSSLLSOVFGKLTASSDSLERNGLQENTACTSAYIOVEPHLLFSSE 946
DB 538 KAKVAFPLDFEFSKSVFLHLQHLHILGAGSDSHEQDSTADONTALLRPHLYEADVLFTRS 597
QY 947 STLHRVEVHPYGTLPV--GPGPEFKTTLRT----- 974
DB 598 SSLSHPEVKANSSLESYDGGIGPPENCVKQNLGFFPIHGVMMKITVPIATRGGNRLML 657
QY 975 -----NNASC-IVQNLTPEPPGVPHPELOHTNRLNGSTQCQVVRCHLGOLAKGTEV 1026
DB 658 RDPFTDQNTSCNIGWNGSTBYRSTPTB-EDLSHAPQRNHSNSDVVSIICNL-RLAPSOET 715
QY 1027 SVGLLRVHNEFFRRARFKSLTVVSTFELGTGTEGSLVQLTEASRWSLSLEVVOTRPLI 1086
DB 716 SFVLVGNLWLTSLKALKYRSKLTIVNALQRPSPPIFREEDPSRQVTFEISKOEDMQV 775
QY 1087 SLWLIGSVLGGLLALLVFLCWLKLGFFAHKK-----IPEEEK 1125
DB 776 PIWIVGSTLGGLLALLVLMWLKLGFFKSAKRREPGLGPIKELK 823

RESULT 14
ITAD RAT STANDARD; PRT; 1161 AA.
AC OQ0YE7;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DE Integrin alpha-D precursor.
GN Name=Itgad;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RA O'Brien M.M., VanderVliet M., Kilgannon P.D., Dietsch G.,
RA Gallatin W.M.;
RT "Cloning of rat alpha D, a novel beta 2 integrin.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Integrin alpha-D/beta-2 is a receptor for ICAM3 and
CC VCAM1. May play a role in the atherosclerotic process such as
CC clearing lipoproteins from plaques and in phagocytosis of blood-
CC borne pathogens, particulate matter, and senescent erythrocytes
CC from the blood. (By similarity).
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-D
CC associates with beta-2 (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC -!- SIMILARITY: Contains 1 VWFA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

-----
ENBL; AF021334; AAF21241.1; -.
DR HSP; P11215; 18HQ.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS00234; VWFA; 1.
KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;
KW Repeat; Signal; Transmembrane.
FT SIGNAL 1 19 Potential.
FT CHAIN 20 1161 Integrin alpha-D.
FT DOMAIN 20 1100 Extracellular (Potential).
FT TRANSMEM 1101 1121 Potential.
FT DOMAIN 1122 1161 Cytoplasmic (Potential).
FT REPEAT 34 87 FG-GAP 1.
FT REPEAT 88 ? FG-GAP 2.
FT DOMAIN 152 334 VWFA.
FT REPEAT 352 402 FG-GAP 3.
FT REPEAT 403 454 FG-GAP 4.
FT REPEAT 456 517 FG-GAP 5.
FT REPEAT 519 577 FG-GAP 6.
FT REPEAT 582 634 FG-GAP 7.
FT CA_BIND 467 475 Potential.
FT CA_BIND 531 539 Potential.
FT CA_BIND 594 602 Potential.
FT SITE 1126 1130 GPFKR motif.
FT DISULFID 69 76 By similarity.
FT DISULFID 108 126 By similarity.
FT DISULFID 656 711 By similarity.
FT DISULFID 769 775 By similarity.
FT DISULFID 845 860 By similarity.
FT DISULFID 993 1017 By similarity.
FT DISULFID 1022 1027 By similarity.
FT CARBOHYD 61 61 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 89 89 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 245 245 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 393 393 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 696 696 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 734 734 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 784 784 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 907 907 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 936 936 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1045 1045 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 1161 AA; 126600 MW; 2258491A984A705E CRC64;

Query Match 19.1%; Score 1135.5; DB 1; Length 1161;
Best Local Similarity 30.3%; Pred. No. 4.2e-72;
Matches 358; Conservative 203; Mismatches 483; Indels 139; Gaps 44;

Qy 20 CSPENLDEHPLRPPGPEAFGVSVLQVGGORWMLVGAPWDPGSDRGDVYRCVPG 79
Db 17 CHGSLNDVEPIVF-REDAASFGTVVQF--GGR-LVVGAPLEAVAVNQTRLYDC--- 69
Qy 80 GAHNAPECAKGLHDYQLGNSHSHAVNNHGLMSLLETDDGGFMACAPLWGRACGSSVFSS 139
Db 70 ----AP-ATGMCQPIVL-RSPLEAVNMSLGLSVLTATNNAQLLACGTAQACVKNMYAK 123
Qy 140 GICARVDASFOQSGSLAPTAQRCTY-MDVVIVLDGNSI--YPMSEVQTFRLRLVGLKF 196
Db 124 GSCLLGSSLQFIQAVPASPMECPQEMDIAFLDGSINQDRFAQWKDFVKALMGE-F 182
Qy 197 IDPQIQVGLVQYGESPVHESLGDFTKEVWRAAKNLSRREGRETKTAQAINVACTEG 256
Db 183 ASTSTL-FSLMQYINILKTHFTTFTEKNILDPQSLVDPIVOLQGL-TYTAGIKTVMEEL 240
Qy 257 FSQSHOGRPEARLLVVVVTGESH-DGEELPAALKACEAGRVTRYGIAGVLGHLRQRDP 315

```

```

Db 241 FHSKNGSRSAKKILLVITDQYKRPLEYSDVI PAADKAGIIRYAIGVDAF---QEP 296
Qy 316 SSFLRIRTIASDPDERFFPNVTDEALTDIVDALGDRIFGLEGSHAENESSGLEMSQI 375
Db 297 TA-LKELNTIGSAPPDQHVFKVGNFAALRSIQRLQOEKIFAIEGTOSRSSSPHMSQE 355
Qy 376 GFSTHRLKDGILFGMGAYDMGGSVLWLEGHRLFPF-----RMALEDEFPALQNH 428
Db 356 GFSSALTSQGVLGAVGSFSW-----SGGAPLYPNTPTPTINMSQEN-----VDMRD 403
Qy 429 AYLGYSVSMLRGGRRRLFSGRFRHRGKVIAFQKKDGA VRAQSLQSGEIGSYFGS 488
Db 404 SYLGSTAVAFWKGVHSLIL-GAPRQHTGKVIF-TQEARHWRPKSEVRGTQIGSYFGA 461
Qy 489 ELCPLDTDRGTTDLVLAAPMFLGPONKETGRVYVVLV-GQOSLLTLQGTLPPEPQD- 546
Db 462 SLCSVDVDRDGSTDLVLIGAPHY--BOTRGQGVSVFPVPGVRGRWQCEATLHGEQHPW 519
Qy 547 ARFGFAMGALPDNLQDGFADVAVGAPLEGDHQAALYLHG-TQSGVRPHPAQRIAAAMP 605
Db 520 GRFGVALTVLGDVNGDNLADVAIGAPGEESRGAVIFHGASRLTMPSPSRVTSQLS 579
Qy 606 HALSYFGRSVDGRDLDDGDDLDVAVGAQGAAILLSRPIVHLTPSLVTPQAISVQRD 665
Db 580 LRLQYFGQSLSGGQDLTQDGLDLAVGAQGHVLLRSLPLLKVELSIRPAPMEVAKAVYQ 639
Qy 666 CRRR-----GOEAVCLTAALCFQVTSRTPGRWDH-QFYMRFTASLDEWTAGARAFDG 717
Db 640 CWERTPTVLEAGBATVCLT-----VHKGSPDLGNVQGSVRYDLALDPLRLSRAIFDE 693
Qy 718 SGQRLSPRLRLSVGNVTCBOLHFHVL-DTSDYLRPVALTPTFAL-DNTTKP---GPVLN 772
Db 694 TKNCTITGRKTLGLD-HCETVKLLIPDCVEDAVSFILILNLSFVDRDSAPRNLRHVL 752
Qy 773 EGSPTSIQKLVPPSKDCGPDNECVTLVLQVNMNDRGSKAPFVVRGGRKVLVSTTLEN 832
Db 753 VGSQDHITASLPFEKCKQELLCEGLGISFNFS-----GLQVLVVGSPDLTVTVTN 807
Qy 833 RKENAYNTSLIIFSRNLHLASLTPORES---PIKVEC-AAPSAHARL-----CSVGHVPV 884
Db 808 EGSDSYGLTVKFYYPAGLSYRRVTGTQOQPHQYFLRLACEABPAQAQEDLRSSSCSINHPIF 867
Qy 885 QTCAKVTFLLEPFSCSSLLSVFGKLTASSDSLERNGTLQENTA-QTSAYIQV----- 937
Db 868 REGAKTTFMITFDVSYKAFGLGDL--LLRAKASSENKPKDPTNKTAFQLELPVKYTVYTLI 925
Qy 938 -----EPHLLPSS-----ESTLHRYEYHPYGTLPVGPGEPEKTTLRN----- 975
Db 926 SRQEDSTNHNVFSSSHGRRQEAHRYRVNLSPLKLA VRNFWPVLNGVAVMDVTLS 985
Qy 976 ----NASCIVONLTETPPGPPVHPEELQHTNRLNGSNTQCVVCHLGOLAKGTEVSGLL 1031
Db 986 SPAQGVSCVSQ--MKPPQNPDLFTQIQRSLVDCSIADCLHFRCDIPSLDIQDELDFILR 1043
Qy 1032 RLVNHEFFRAKPSKLTVTSTPELGTGEGSVLQI--TEA---SRWSSLELVQVTRPILI 1086
Db 1044 GNLSFGWSTQLEKULLVSEAITDTSVYSQLPQEAFLRAQVETLSEYVYVEFI-- 1101
Qy 1087 SLMLIGSVLGGLLALLVLCMLKGLFFAHKKIPBEEREK 1129
Db 1102 --FLVAGSSVGGLLALLVTVLYLKLGF-----KQYK 1133

RESULT 15
ITAD HUMAN
ID ITAD HUMAN STANDARD; PRT; 1162 AA.
AC Q13349; Q15575; Q15576;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-D precursor (leukointegrin alpha D) (CD11d) (ADB2).
GN Name=ITGAD;

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 143.162 Seconds  
(without alignments)  
3058.161 Million cell updates/sec

Title: US-09-647-544-4  
Perfect score: 5932  
Sequence: 1 MEPLFVTHLPFLVLTGLC.....GFFAHKKIPREKREKLEQ 1132

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq16Dec04:.\*  
1: Geneseqp1980s:.\*  
2: Geneseqp1990s:.\*  
3: Geneseqp2000s:.\*  
4: Geneseqp2001s:.\*  
5: Geneseqp2002s:.\*  
6: Geneseqp2003as:.\*  
7: Geneseqp2003bs:.\*  
8: Geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5932	100.0	1132	3 AAY32243	Aay32243 Human int
2	5904.5	99.5	1167	3 AAY32242	Aay32242 Human int
3	5890.5	99.3	1167	8 ADQ19290	Adq19290 Human sof
4	5838.5	98.4	1167	4 AAB64584	Aab64584 Human sec
5	5838.5	98.4	1167	6 ABP99490	Abp99490 Human sec
6	5838.5	98.4	1167	6 ABR00964	Abr00964 Human gen
7	5838.5	98.4	1167	6 ADA44026	Ada44026 Human sec
8	5829.5	98.3	1152	4 AAB64657	Aab64657 Human sec
9	5763.5	97.2	1152	4 AAB64658	Aab64658 Human sec
10	4878.5	82.2	1049	5 ADR41424	Adr41424 Human CD-
11	2355	39.7	1188	4 AAU14231	Aau14231 Human nov
12	2355	39.7	1188	4 AAB50085	Aab50085 Human A25
13	2355	39.7	1188	5 AAU10551	Aau10551 Human A25
14	2354	39.7	1188	4 AAU14467	Aau14467 Human nov
15	2354	39.7	1188	7 ADE09956	Ade09956 Novel pro
16	2344.5	39.5	1189	3 AAB25582	Aab25582 ITGA11 pr
17	2344.5	39.5	1189	4 ABG12949	Abg12949 Novel hum
18	2344.5	39.5	1189	6 ABR58364	Abr58364 Human NOV
19	2344.5	39.5	1189	6 ADA27054	Ada27054 Human nov
20	2344.5	39.5	1189	7 ADE63570	Ade63570 Human Pro
21	2344.5	39.5	1189	8 ADE86584	Ade86584 Novel hum
22	2344.5	39.5	1189	8 ADQ19968	Adq19968 Human sof
23	2337.5	39.4	1188	4 AAB50087	Aab50087 Murine A2
24	2337.5	39.4	1188	5 AAU10552	Aau10552 Murine A2
25	2319	39.1	1188	4 AAB30929	Aab30929 Amino aci

26	2269.5	38.3	1189	8 ADH80785	Adh80785 Human pol
27	2194	37.0	1034	3 AAB25590	Aab25590 Protein e
28	2194	37.0	1034	6 ADA27062	Ada27062 Human nov
29	2194	37.0	1034	8 ADE86592	Ade86592 Novel hum
30	2194	37.0	1058	5 ADR41496	Adr41496 Human CD-
31	2140	36.1	1120	6 ABR58365	Abr58365 Human NOV
32	1979	33.4	987	8 ADP29492	Adp29492 Human sec
33	1887.5	31.8	1179	5 ABB90759	Abb90759 Human Tum
34	1887.5	31.8	1179	5 ABP64915	Abp64915 Human pro
35	1887.5	31.8	1179	6 ABU54466	Abu54466 Human tum
36	1884	31.8	1180	5 ABB90788	Abb90788 Rat Tumou
37	1884	31.8	1180	6 ABUS4495	Abu54495 Mouse tum
38	1884	31.8	1180	7 ADE63568	Ade63568 Rat Prote
39	1881.5	31.7	1151	8 ADE86652	Ade86652 Human int
40	1874.5	31.6	1177	8 ADQ39498	Adq39498 Human myo
41	1825.5	30.8	1183	2 AAY07728	Aay07728 Armenian
42	1733.5	29.2	1181	6 ABU03548	Abu03548 Angiogene
43	1733.5	29.2	1181	6 ABR59703	Abr59703 Human VLA
44	1733.5	29.2	1181	6 AAG79775	Aag79775 Alpha2 in
45	1733.5	29.2	1181	6 ABU03616	Abu03616 Human exp

ALIGNMENTS

RESULT 1

AAAY32243  
ID AAAY32243 standard; protein; 1132 AA.

XX AC AAY32243;  
XX AC AAY32243;  
DT 15-FEB-2000 (first entry)  
XX DE Human integrin subunit alpha-10 splice variant.

XX KW Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;  
XX KW osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;  
XX KW therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker;  
XX KW splice variant.

XX OS Homo sapiens.

XX FH Location/Qualifiers  
FT Peptide i. .22  
FT Protein /note= "signal peptide"  
FT FT 23. .1132  
FT FT /note= "mature protein"

XX WO9951639-A1.

XX PD 14-OCT-1999.

XX PF 31-MAR-1999; 99WO-SE000544.

XX PR 02-APR-1998; 98SE-00001164.

XX PR 28-JAN-1999; 99SE-00000319.

XX PA (ACTI-) ACTIVE BIOTECH AB.

XX PI Lundgren-Akerlund E;

XX WPI; 2000-052639/04.

XX DR N-ESDB; AAZ34720.

XX PT New isolated integrin subunit alpha-10, used as a marker or target molecule for cells during development, regeneration and pathological conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or inflammation.

XX PS Claim 1; Page 43-48; 90pp; English.

XX CC This sequence represents a splice variant of novel human chondrocyte integrin subunit alpha-10 (ISa10). It is identical to ISa10 (see









```

Db      1141 VFCLWKLGFPAHKIPBEKREKLEQ 1167
RESULT 3
ADQ19290
ID      ADQ19290 standard; protein; 1167 AA.
XX      AC
XX      ADQ19290;
XX      DT
XX      26-AUG-2004 (first entry)
XX      DE
XX      Human soft tissue sarcoma-upregulated protein - SEQ ID 2109.
XX      soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
XX      KW
XX      Homo sapiens.
XX      OS
XX      WO2004048938-A2.
XX      PN
XX      10-JUN-2004.
XX      PD
XX      26-NOV-2003; 2003WO-US038193.
XX      PF
XX      26-NOV-2002; 2002US-0429739P.
XX      PR
XX      (PROT-) PROTEIN DESIGN LABS INC.
XX      PA
XX      Aziz N, Gineburg WM, Zlotnik A;
XX      PI
XX      WPI; 2004-441208/41.
XX      DR
XX      Early detection of soft tissue sarcoma comprises determining expression
XX      of a gene in a first soft tissue sample and a normal soft tissue sample
XX      PT      and comparing the gene expression, also useful in treating soft tissue
XX      PT      sarcoma.
XX      PS
XX      Example 2; SEQ ID NO 2109; 210pp; English.
XX      CC
XX      The invention relates to a novel method for detecting soft tissue sarcoma
XX      which comprises obtaining a first soft tissue sample from an individual
XX      CC      and a normal soft tissue sample from the same or different individual,
XX      CC      determining the expression of a gene in both samples and comparing the
XX      CC      expression of the gene in both soft tissue samples, where a higher level
XX      CC      of protein expression in the first soft tissue sample indicates the
XX      CC      presence of soft tissue sarcoma. The method of the invention has
XX      CC      cytosstatic applications and may be useful for detecting soft tissue
XX      CC      sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX      CC      acid sequences may be useful in diagnostic and screening applications.
XX      CC      The current sequence is that of a human soft tissue sarcoma-upregulated
XX      CC      protein of the invention. The current sequence is not shown within the
XX      CC      specification per se but was submitted in CD format by the inventor.
XX      SQ
XX      Sequence 1167 AA;

Query Match      99.3%; Score 5890.5; DB 8; Length 1167;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1129; Conservative 2; Mismatches 1; Indels 35; Gaps 1;

Qy      1 MELPFTVHLFLPLVFLTGLCSPPNLDHHPRLFPGPPEAREFGYVLQHVGGQRMVLGA 60
Db      1 MELPFTVHLFLPLVFLTGLCSPPNLDHHPRLFPGPPEAREFGYVLQHVGGQRMVLGA 60
Qy      61 PWDGPGDRRGDVRPCVGGAHNAPCAKGLHGDYQLGNSSHPAVNMHLGMSLLTDDGGG 120
Db      61 PWDGPGDRRGDVRPCVGGAHNAPCAKGLHGDYQLGNSSHPAVNMHLGMSLLTDDGGG 120
Qy      121 FMACAPLWSRACGSSVFSSGICARVDASFPQGS LAPTAQRCTPTMDVIVLDGNSIYP 180
Db      121 FMACAPLWSRACGSSVFSSGICARVDASFPQGS LAPTAQRCTPTMDVIVLDGNSIYP 180
Qy      181 WSEVQTFLLRLVGLKFLDIDPEIQVGLVQYGSPPVHWSLGDFTKKEEVRAAKNLSRREG 240
Db      181 WSEVQTFLLRLVGLKFLDIDPEIQVGLVQYGSPPVHWSLGDFTKKEEVRAAKNLSRREG 240

241 RETKTAQAIWVACTEGFSQSHGGRPEAARLLVVVTDGESHGDELPALAKACAGRVTRY 300
241 RETKTAQAIWVACTEGFSQSHGGRPEAARLLVVVTDGESHGDELPALAKACAGRVTRY 300
301 GIAVLGHYLRQRDPSSFLREINTIASDDPDERFFNVNTEAALTIDVLDGDRIFGLEGS 360
301 GIAVLGHYLRQRDPSSFLREINTIASDDPDERFFNVNTEAALTIDVLDGDRIFGLEGS 360
361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGAYDMGGSVLMLEGGHRLLPPPMALDEDEF 420
361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGAYDMGGSVLMLEGGHRLLPPPMALDEDEF 420
421 PPALQNHAAVLYGYSVSSMLRGGRRLLFSGAPRFRHKGKVIAPQLKKDGAVRVAQSLQGE 480
421 PPALQNHAAVLYGYSVSSMLRGGRRLLFSGAPRFRHKGKVIAPQLKKDGAVRVAQSLQGE 480
481 QIGSYFGSELCPDTRDGTVDVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLOQTLLQ 540
481 QIGSYFGSELCPDTRDGTVDVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLOQTLLQ 540
541 PEPPQDARFGFAMGALPDNLNODGFADVAVGAPLEDGHQGALYLYHGTQSGVRPHPAQRIA 600
541 PEPPQDARFGFAMGALPDNLNODGFADVAVGAPLEDGHQGALYLYHGTQSGVRPHPAQRIA 600
601 AASMPHALS YFGRSVDGRDLDDGDDLVDAVAGAGAAIILSSRPVHLTPSLEVTPOAIS 660
601 AASMPHALS YFGRSVDGRDLDDGDDLVDAVAGAGAAIILSSRPVHLTPSLEVTPOAIS 660
661 VVQDCCRROQBAVCLTAALCFQVTSRTPGRWDPHQYMFRTASLDDEWTAGARAAFDGSGQ 720
661 VVQDCCRROQBAVCLTAALCFQVTSRTPGRWDPHQYMFRTASLDDEWTAGARAAFDGSGQ 720
721 RLSPRRLRLSVGNVTCBOLHFHVLDTSDYLRPVALTVPFALDNTTKPGPVNLNMGSPSTIQ 780
721 RLSPRRLRLSVGNVTCBOLHFHVLDTSDYLRPVALTVPFALDNTTKPGPVNLNMGSPSTIQ 780
781 KLVFPKDCGPDNECVTDLVQVNMDIRGSRKAPFVVRGGRKVLVSTTLLENKENAYNT 840
781 KLVFPKDCGPDNECVTDLVQVNMDIRGSRKAPFVVRGGRKVLVSTTLLENKENAYNT 840
841 SLSLIIFSRNLHLASLTQRESPIKVECAAPSAHARLCSVGHVPVOTGAKVTFLLPEFPSC 900
841 SLSLIIFSRNLHLASLTQRESPIKVECAAPSAHARLCSVGHVPVOTGAKVTFLLPEFPSC 900
901 SSLLSQVFGKLTASSDSLEBNGTLQNTAQTSAIOVEPHLLFSSSTLHRYEHPVGTLL 960
901 SSLLSQVFGKLTASSDSLEBNGTLQNTAQTSAIOVEPHLLFSSSTLHRYEHPVGTLL 960
961 PVGPGPEFKTTLR-----TNNASCIQVQLT 985
961 PVGPGPEFKTTLR-----TNNASCIQVQLT 985
986 EPPGPPVHPELOHTNRLNGSNSTCCQVVRCHLQOLAKGTSEVSGLLRLVNEFFRRKAPK 1045
1021 EPPGPPVHPELOHTNRLNGSNSTCCQVVRCHLQOLAKGTSEVSGLLRLVNEFFRRKAPK 1080
1046 SLTVVSTFELGTSEGSVLQLTSEARWSESLLEVQTRPILISWILIGSVLGLLALL 1105
1081 SLTVVSTFELGTSEGSVLQLTSEARWSESLLEVQTRPILISWILIGSVLGLLALL 1140
1106 VFCLWKLGFPAHKIPBEKREKLEQ 1132
1141 VFCLWKLGFPAHKIPBEKREKLEQ 1167

RESULT 4
AAB64584
ID      AAB64584 standard; protein; 1167 AA.
XX      AC
XX      AAB64584;
XX      DT      22-MAR-2001 (first entry)

```

XX Human secreted protein #37.  
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
 KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.  
 XX Homo sapiens.  
 OS  
 XX  
 XX WO200077197-A1.  
 XX  
 XX 21-DEC-2000.  
 XX  
 XX 01-JUN-2000; 2000MO-US014934.  
 XX  
 XX 11-JUN-1999; 99US-0138599P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX (ROSE/) ROSEN C A.  
 XX  
 XX Rosen CA, Ruben SM, Komatsoulis GA;  
 XX  
 XX WPI; 2001-032312/04.  
 XX N-PSDB; AAF32793.  
 XX  
 XX Isolated nucleic acid molecule encoding a human secreted protein is used  
 XX in preventing, treating or ameliorating a medical condition.  
 XX  
 XX Claim 11; Page 496-500; 558pp; English.  
 XX  
 XX Sequences AAB64549-B64594 represent the amino acid sequences of 47 human  
 XX secreted proteins encoded by the genes AAF32757-F32803. The genes and  
 XX proteins are useful for preventing, ameliorating or treating medical  
 XX conditions, e.g. by protein or gene therapy. The genes are isolated from  
 XX a range of human tissues disclosed in the specification. The nucleic  
 XX acids, proteins, antibodies and (ant)agonists are useful in the  
 XX diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
 XX ovarian cancer, and other cancers of the adrenal gland, bone, bone  
 XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b)  
 XX immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
 XX anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
 XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
 XX cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
 XX ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
 XX infectious diseases such as viral, bacterial, fungal and parasitic  
 XX infections  
 XX  
 XX Query Match 98.4%; Score 5838.5; DB 4; Length 1167;  
 XX Best Local Similarity 96.0%; Pred. No. 0;  
 XX Matches 1120; Conservative 3; Mismatches 9; Indels 35; Gaps 1;  
 XX  
 XX 1 MELPFTVTHLFLVFLTLGLCSFNLDEHHPLRFPGPPEAFEGYVLQHVGGQRWMLVGA 60  
 XX :  
 XX 1 VELPFTVTHLFLVFLTLGLCSFNLDEHHPLRFPGPPEAFEGYVLQHVGGQRWMLVGA 60  
 XX  
 XX 61 PWDGSGDRGDRVRCVGGAHNAPCAKGLGDLQNGNSHPANVMHGLMSLLTDDGG 120  
 XX :  
 XX 61 PWDGSGDRGDRVRCVGGAHNAPCAKGLGDLQNGNSHPANVMHGLMSLLTDDGG 120  
 XX  
 XX 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPYMDVIVLDGNSNIYP 180  
 XX :  
 XX 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPYMDVIVLDGNSNIYP 180  
 XX  
 XX 181 WSEVOTFLRLRVGLKFLIDPEIQVGLVQYGSPPVHWSLGDIFRTKEEVVRAAKNLSREG 240  
 XX :  
 XX 181 WSEVOTFLRLRVGLKFLIDPEIQVGLVQYGSPPVHWSLGDIFRTKEEVVRAAKNLSREG 240  
 XX  
 XX 241 RETKTAQIMVACTEGFSQSHGGRPEAARLLVVVDGSHDGBELPAALKACEAGRVTY 300  
 XX :  
 XX 241 RETKTAQIMVACTEGFSQSHGGRPEAARLLVVVDGSHDGBELPAALKACEAGRVTY 300

Db 241 RETKTAQIMVACTEGFSQSHGGRPEAARLLVVVDGSHDGBELPAALKACEAGRVTY 300  
 Qy 301 GIAVLGHLRQRDPSSFLREIRTIASDPDERFFNTVDEAALTDIYDALGDRIFGLEGS 360  
 Db 301 GIAVLGHLRQRDPSSFLREIRTIASDPDERFFNTVDEAALTDIYDALGDRIFGLEGS 360  
 Qy 361 HAENESSFGLEMSQIGFSTHRLKDGILFGMVGAYDMGGSVLWLEGHRLPPRMALEDEF 420  
 Db 361 HAENESSFGLEMSQIGFSTHRLKDGILFGMVGAYDMGGSVLWLEGHRLPPRMALEDEF 420  
 Qy 421 PPALQNHAAAYLGYSVSSMLLRGGRRLFLSGAPFRHRGKVIATFOLKKDGAVRVAQSLQGE 480  
 Db 421 PPALQNHAAAYLGYSVSSMLLRGGRRLFLSGAPFRHRGKVIATFOLKKDGAVRVAQSLQGE 480  
 Qy 481 QIGSYFSGELCPDLTDRDGTDLVLLVAAPFLGPONKETGRVYVYLQQOSSLTLQGTLO 540  
 Db 481 QIGSYFSGELCPDLTDRDGTDLVLLVAAPFLGPONKETGRVYVYLQQOSSLTLQGTLO 540  
 Qy 541 PEPQDARFGFAMGALPDLNODGFADVAVGAPLEDGHQGALYLYHGTQSGVRPAPORIA 600  
 Db 541 PEPQDARFGFAMGALPDLNODGFADVAVGAPLEDGHQGALYLYHGTQSGVRPAPORIA 600  
 Qy 601 AASMPHALSYFGRSVDGRLDLGDGLVDVAVGAQAAAILLSSRPVHLTPSLEVTPOAIS 660  
 Db 601 AASMPHALSYFGRSVDGRLDLGDGLVDVAVGAQAAAILLSSRPVHLTPSLEVTPOAIS 660  
 Qy 661 VVQDCRRRGOEAVCLTAALCFQVTSRTPGRDHGFYMRFTASLDEWTAGARAAFDGSGQ 720  
 Db 661 VVQDCRRRGOEAVCLTAALCFQVTSRTPGRDHGFYMRFTASLDEWTAGARAAFDGSGQ 720  
 Qy 721 RLSRRRLSVGNVTCQLHFLVLDTSYLPRPVALTVTTEALDNTTKPGVNLNCSPTSIO 780  
 Db 721 RLSRRRLSVGNVTCQLHFLVLDTSYLPRPVALTVTTEALDNTTKPGVNLNCSPTSIO 780  
 Qy 781 KLVPPFSKDCPDNECVTDVLQVNMDIRGSRKAPFVVRGRRKVLVSTTLNKRKENAYNT 840  
 Db 781 KLVPPFSKDCPDNECVTDVLQVNMDIRGSRKAPFVVRGRRKVLVSTTLNKRKENAYNT 840  
 Qy 841 SLSIFSRNLHLASLTLPQRESPIKVECAAPSAHARLCSVGHVPFQTKAKVTLLEFEFSC 900  
 Db 841 SLSIFSRNLHLASLTLPQRESPIKVECAAPSAHARLCSVGHVPFQTKAKVTLLEFEFSC 900  
 Qy 901 SLSLQVFGKLTASSDSLERNGLTOENTACTSAIOYEPHLLFSSESTLHRYEYVHPYGL 960  
 Db 901 SLSLQVFGKLTASSDSLERNGLTOENTACTSAIOYEPHLLFSSESTLHRYEYVHPYGL 960  
 Qy 961 PVGPGPEFKTTLR-----TNNASCIQVNLIT 985  
 Db 961 PVGPGPEFKTTLR-----TNNASCIQVNLIT 985  
 Qy 986 EPPGPPVHPELOHTNRLNGSNTQCQVVRCHLQAKGTEVSVGLLRVHNEFFRRAKFK 1045  
 Db 1021 EPPGPPVHPELOHTNRLNGSNTQCQVVRCHLQAKGTEVSVGLLRVHNEFFRRAKFK 1080  
 Qy 1046 SLTVSVTFELGTEGSLVQLTEASRWSESLLEVVTQRPILISLWILGSLGGLLALL 1105  
 Db 1081 SLTVSVTFELGTEGSLVQLTEASRWSESLLEVVTQRPILISLWILGSLGGLLALL 1140  
 Qy 1106 VFCLWLKGLFFAHKKIPBEEKREKLEQ 1132  
 Db 1141 VFCLWLKGLFFAHKKIPBEEKREKLEQ 1167  
 XX  
 XX RESULT 5  
 XX ABP99490  
 XX ID ABP99490 standard; protein; 1167 AA.  
 XX AC ABP99490;  
 XX DT 26-MAR-2003 (first entry)  
 XX DE Human secreted protein SEQ ID NO 434.  
 XX

Human; secreted protein; nontropic; neuroprotective; cytostatic; virucide; dermatological; immunosuppressive; anti-inflammatory; anti-viral; antiparasitic; antiparkinsonian; antitickling; antianemic; antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective; antiinflammatory; antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine; cardiovascular disorder; neurological disease; nephrotropic; gene therapy.

OS Homo sapiens.

XX W0200277186-A2.

XX 03-OCT-2002.

XX 26-MAR-2002; 2002WO-US009188.

XX 27-MAR-2001; 2001US-0278650P.

XX 12-SEP-2001; 2001US-00950082.

XX 12-SEP-2001; 2001US-00950083.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2003-040583/03.

XX N-PSDB; ABZ66911.

XX New human secreted proteins encoded by genes contained in cDNA clones (e.g. HGAC19), useful for preventing, treating or diagnosing e.g. AIDS, multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or West Nile fever.

XX Claim 1; Page 1395-1398; 2423pp; English.

XX The invention relates to novel human genes (ABZ66891-ABZ68209) and the encoded secreted proteins (ABP99470-ABP99872) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections

XX Sequence 1167 AA;

Query Match 98.4%; Score 5838.5; DB 6; Length 1167;

Best Local Similarity 96.0%; Pred. No. 0;

Matches 1120; Conservative 3; Mismatches 9; Indels 35; Gaps 1;

QY 1 MELPFVTHLFLPLVFLTGLCSFNLDEHRLFPPEAEFGYSLVHVGQGRWMLVGA 60

DB 1 VELPFVTHLFLPLVFLTGLCSFNLDEHRLFPPEAEFGYSLVHVGQGRWMLVGA 60

QY 61 PWDPGSDRRGDVYRCPVGAHNAAPCAKHLGDYQLGNSHPAVNMHLGMSLLETDDGG 120

DB 61 PWDPGSDRRGDVYRCPVGAHNAAPCAKHLGDYQLGNSHPAVNMHLGMSLLETDDGG 120

QY 121 FMACAPLWSRACGSSVSSGICARVDASFPQGS LAPTAQRCPYMDVIVLDGNSIYP 180

DB 121 FMACAPLWSRACGSSVSSGICARVDASFPQGS LAPTAQRCPYMDVIVLDGNSIYP 180

QY 181 WSEVQTFLRLVGLFDIDPQIQVLGVQGESPVHWSLGDFTKEEVRAAKNLSREG 240

DB 181 WSEVQTFLRLVGLFDIDPQIQVLGVQGESPVHWSLGDFTKEEVRAAKNLSREG 240

QY 241 RETKTAQAINVACTEGFSQSHGQRPEAARLLVVVTGDESHGDELPALAKACAGRVTRY 300

DB 241 RETKTAQAINVACTEGFSQSHGQRPEAARLLVVVTGDESHGDELPALAKACAGRVTRY 300

QY 301 GIAVLGHYLRQRDPSPFLREIRTIASDPDERFFPNVTDEAALTDIVDALGDRIFGLEGS 360

DB 301 GIAVLGHYLRQRDPSPFLREIRTIASDPDERFFPNVTDEAALTDIVDALGDRIFGLEGS 360

QY 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGAYDMGGSVLMLEGHRLPPPMALDEDEF 420

DB 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGAYDMGGSVLMLEGHRLPPPMALDEDEF 420

QY 421 PPALQNHAAAYLYGYSVSMLLRGRRFLSGAPFRHRGKVIAFQLKKOGAVRVAQSLQGE 480

DB 421 PPALQNHAAAYLYGYSVSMLLRGRRFLSGAPFRHRGKVIAFQLKKOGAVRVAQSLQGE 480

QY 481 QIGSYFGSELCLDTRDGTDDVLLVAAPMFLGPONKGTGRVVVYLVGQSSLLTLQSTLQ 540

DB 481 QIGSYFGSELCLDTRDGTDDVLLVAAPMFLGPONKGTGRVVVYLVGQSSLLTLQSTLQ 540

QY 541 PEPPQDARFGFAMGALPDNLNODGFADVAVGAPLEDGHQALYLYHGTQSGVRPHPAQRIA 600

DB 541 PEPPQDARFGFAMGALPDNLNODGFADVAVGAPLEDGHQALYLYHGTQSGVRPHPAQRIA 600

QY 601 AASMPHALSYFGRSVDRLDGDLDVAVAGAGAAILLSSRPPIVHLTSPSLEVTPOAIS 660

DB 601 AASMPHALSYFGRSVDRLDGDLDVAVAGAGAAILLSSRPPIVHLTSPSLEVTPOAIS 660

QY 661 VWORDCRRQGEAVCLTAALCFQVTSRTQGRWHDQFYMRTASLDEWTAGARAAFDGSGQ 720

DB 661 VWORDCRRQGEAVCLTAALCFQVTSRTQGRWHDQFYMRTASLDEWTAGARAAFDGSGQ 720

QY 721 RLSPRRLRLSVGNVTCQLHFFHVLDTSDYLRPVALTVPFALDNTTKPGPVNLGSGSPSIQ 780

DB 721 RLSPRRLRLSVGNVTCQLHFFHVLDTSDYLRPVALTVPFALDNTTKPGPVNLGSGSPSIQ 780

QY 781 KLVFPKDCGPDNECVTDLVQVNMDIRSKAPFVVRGRRKVLVSTTLLENKENAYNT 840

DB 781 KLVFPKDCGPDNECVTDLVQVNMDIRSKAPFVVRGRRKVLVSTTLLENKENAYNT 840

QY 841 SLSIIISRNHLASLPQRESPIKVECAAPSAHARLCSVGHVPVOTGAKVTFLLEPFSC 900

DB 841 SLSIIISRNHLASLPQRESPIKVECAAPSAHARLCSVGHVPVOTGAKVTFLLEPFSC 900

QY 901 SSLLSVOFGKLTASSDSLRNGTLQNTAQTSAIYQEPHLLFSSSTLHRYEHPVGTLL 960

DB 901 SSLLSVOFGKLTASSDSLRNGTLQNTAQTSAIYQEPHLLFSSSTLHRYEHPVGTLL 960

QY 961 PVGPGPEFKTTLR-----TNNASCIQVQLT 985

DB 961 PVGPGPEFKTTLR-----TNNASCIQVQLT 985

QY 986 EPGPPVHPPELOHTNRLNGSNTCCQVVRCHLQLAGTSEVSGLLRLVNEFFRRKAPFK 1045

DB 1021 EPGPPVHPPELOHTNRLNGSNTCCQVVRCHLQLAGTSEVSGLLRLVNEFFRRKAPFK 1080

QY 1046 SLTVVSTFELGTSEGSVLQLEASRWSESLEVVQTRPILISWILLIGSVLGLLLALL 1105

DB 1081 SLTVVSTFELGTSEGSVLQLEASRWSESLEVVQTRPILISWILLIGSVLGLLLALL 1140

QY 1106 VFCLWKLGFPAHKKIPPEEKREKLEQ 1132

DB 1141 VFCLWKLGFPAHKKIPPEEKREKLEQ 1167

RESULT 6

ABR00964

ID ABR00964 standard; protein; 1167 AA.

XX AC ABR00964;

XX DT 12-MAY-2003 (first entry)

XX



XX ADA44026;  
XX AC  
XX 20-NOV-2003 (first entry)  
XX DE Human secreted protein SEQ ID 214.  
XX KW Gene therapy; human; Antidiabetic; Anorectic; Ophthalmological;  
XX KW Neuroprotective; Cerebroprotective; Antianemic.  
XX OS Homo sapiens.  
XX PN WO2003000865-A2.  
XX PD 03-JAN-2003.  
XX PF 26-MAR-2002; 2002WO-US009105.  
XX PR 27-MAR-2001; 2001US-0278650P.  
XX PR 12-SEP-2001; 2001US-00950082.  
XX PR 12-SEP-2001; 2001US-00950083.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Ruben SM;  
XX DR WPI; 2003-184045/18.  
XX DR N-PSDB; ADA43832.  
XX PT A human secreted protein and nucleic acids useful for preparing a  
XX PT diagnostic or pharmaceutical composition for diagnosing or treating  
XX PT diabetes or conditions related to diabetes, e.g. hyperglycemia, obesity,  
XX PT retinopathy, neuropathy.  
XX PS Claim 1; SEQ ID NO 214; 701pp; English.  
XX CC The invention relates to novel genes and their fragments which are useful  
XX CC for preventing, treating or ameliorating medical conditions e.g. by  
XX CC protein or gene therapy. The genes are isolated from a range of human  
XX CC tissues disclosed in the specification. The nucleic acids and proteins  
XX CC are useful in the diagnosis, treatment and prevention of conditions  
XX CC related to diabetes, e.g. hyperglycaemia, obesity, retinopathy,  
XX CC polynuropathy, atherosclerosis, anaemia, stroke, gangrene, impotence,  
XX CC infection, cataract, renal disorders, or endocrine disorders. The present  
XX CC sequence was used to illustrate the invention.  
XX SQ Sequence 1167 AA;  
Query Match 98.4%; Score 5838.5; DB 6; Length 1167;  
Best Local Similarity 96.0%; Pred. No. 0;  
Matches 1120; Conservative 3; Mismatches 9; Indels 35; Gaps 1;  
Qy 1 MELDFVTHLFLPLVFLTGLCSPPNLDHHPRLFPGPPEAEFGYSLQHVGGQRWMLVGA 60  
Dy :  
Dy 1 VELPFVTHLFLPLVFLTGLCSPPNLDHHPRLFPGPPEAEFGYSLQHVGGQRWMLVGA 60  
Qy 61 PWDGPGDRRGDVRCPVGGAHNAPCAKHGLDYLQGNSSHPANMHLGMSLLETDDGG 120  
Dy 61 PWDGPGDRRGDVRCPVGGAHNAPCAKHGLDYLQGNSSHPANMHLGMSLLETDDGG 120  
Qy 121 FMACAPLMSRACGSSVFSSGICARVDASFPQGS LAPTAQRCPTMYDVIVLDGNSIYP 180  
Dy 121 FMACAPLMSRACGSSVFSSGICARVDASFPQGS LAPTAQRCPTMYDVIVLDGNSIYP 180  
Qy 181 WSEVQTFELRLVGLKFLTDPEIQVGLVQYGSPPHWSLGDPRFKEEVVRAAKNLSRREG 240  
Dy 181 WSEVQTFELRLVGLKFLTDPEIQVGLVQYGSPPHWSLGDPRFKEEVVRAAKNLSRREG 240  
Qy 241 RETTQAQIMVACTEGFSQSHGGRPEARLAVVTDGSHDGEELPAALKAACEAGRVTRY 300  
Dy 241 RETTQAQIMVACTEGFSQSHGGRPEARLAVVTDGSHDGEELPAALKAACEAGRVTRY 300  
Qy 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNVTDAAALTDIVDALGDRIFGLEGS 360

Db 301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNVTDAAALTDIVDALGDRIFGLEGS 360  
Qy 361 HAENESSFGLMSQIGFSTHRLKXGILFGMVAGYDMGGSVLWLEGHRLPPPPMALEDEF 420  
Db 361 HAENESSFGLMSQIGFSTHRLKXGILFGMVAGYDMGGSVLWLEGHRLPPPPMALEDEF 420  
Qy 421 PPALQNHAAVLYGYSVSMILLRGRRRLPLSGAPFRHRGKVIAPQLKXGAVRVAQSLQGE 480  
Db 421 PPALQNHAAVLYGYSVSMILLRGRRRLPLSGAPFRHRGKVIAPQLKXGAVRVAQSLQGE 480  
Qy 481 QIGSYFSGELCPLDTRDGTDLVLAAPMFLGPQNKETGRVYVYLVGQOSLTLQGTQL 540  
Db 481 QIGSYFSGELCPLDTRDGTDLVLAAPMFLGPQNKETGRVYVYLVGQOSLTLQGTQL 540  
Qy 541 PEPQDARFGFANGALPDNLNODGFADVAVGAPLEDHGGALYLYHGTQSGVRPAPQRIA 600  
Db 541 PEPQDARFGFANGALPDNLNODGFADVAVGAPLEDHGGALYLYHGTQSGVRPAPQRIA 600  
Qy 601 AASMPHALSYFGRSVDGRDLDDGDDLDVAVAGAAIILLSSRPVHLTPSLEVTQAIS 660  
Db 601 AASMPHALSYFGRSVDGRDLDDGDDLDVAVAGAAIILLSSRPVHLTPSLEVTQAIS 660  
Qy 661 VQDCCRQGEAVCLTAALCFQVTSRTPGRWDPHFYMRFTASLDEWTAGARAAFDGSGQ 720  
Db 661 VQDCCRQGEAVCLTAALCFQVTSRTPGRWDPHFYMRFTASLDEWTAGARAAFDGSGQ 720  
Qy 721 RLSPRRLRLSVGNVTCQLHFFHVLDTSDYLRPVALTVTTFALDNTTKPGPVLNBSGSPSIQ 780  
Db 721 RLSPRRLRLSVGNVTCQLHFFHVLDTSDYLRPVALTVTTFALDNTTKPGPVLNBSGSPSIQ 780  
Qy 781 KLVPFSDKDCPDNECVTDLVQNMIDIRGSRKAPFVVRGRRKVLVSTTLENKENAYNT 840  
Db 781 KLVPFSDKDCPDNECVTDLVQNMIDIRGSRKAPFVVRGRRKVLVSTTLENKENAYNT 840  
Qy 841 SLSIIFSRNLHLASLTLPQRESPIKVECAAPSAHARLCSVGHVPVOTGAKVTFLEPEFSC 900  
Db 841 SLSIIFSRNLHLASLTLPQRESPIKVECAAPSAHARLCSVGHVPVOTGAKVTFLEPEFSC 900  
Qy 901 SSLLSQVFGKLTASSSLERNGLTQNTAQTSAIYQEPHLLFSSSTLHRYEVHPYGTIL 960  
Db 901 SSLLSQVFGKLTASSSLERNGLTQNTAQTSAIYQEPHLLFSSSTLHRYEVHPYGTIL 960  
Qy 961 PVGPGPEFKTTLR-----TNNASCIVQNLT 985  
Db 961 PVGPGPEFKTTLR-----TNNASCIVQNLT 985  
Qy 986 EPPGPPVHPEELQHTNRLNGSNTCCQVVRCHLQGLAKGTSEVSGLLRLVHNEFFRRAKFK 1045  
Db 1021 EPPGPPVHPEELQHTNRLNGSNTCCQVVRCHLQGLAKGTSEVSGLLRLVHNEFFRRAKFK 1080  
Qy 1046 SLTVSTVFELGTREGSVLQLTSEARNSSESLLEVVQTRPILISLWILIGSVLGGILLALL 1105  
Db 1081 SLTVSTVFELGTREGSVLQLTSEARNSSESLLEVVQTRPILISLWILIGSVLGGILLALL 1140  
Qy 1106 VFCLMWKLGFFAHKKIPEEKREKLEQ 1132  
Db 1141 VFCLMWKLGFFAHKKIPEEKREKLEQ 1167  
RESULT 8  
AAB64657  
ID AAB64657 standard; protein; 1152 AA.  
XX AAB64657;  
XX 22-MAR-2001 (first entry)  
XX Human secreted protein BLAST search protein SEQ ID NO: 167.  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
XX antiallergic; hepatocrotic; antiinflammatory; antitumor;  
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;



KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein.  
XX Homo sapiens.  
XX WO200077197-A1.  
XX 21-DEC-2000.  
XX 01-JUN-2000; 2000MO-US014934.  
XX 11-JUN-1999; 99US-0138599P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX (ROSE/) ROSEN C A.  
XX Rosen CA, Ruben SM, Komatsoulis GA;  
XX WPI; 2001-032312/04.  
XX Isolated nucleic acid molecule encoding a human secreted protein is used  
PT in preventing, treating or ameliorating a medical condition.  
XX  
XX Disclosure; Page 543-546; 558pp; English.  
XX The invention relates to the isolation of genes AAF32757-F32803 encoding  
CC the human secreted proteins AAB64549-B64594. The sequence is a search  
CC result from a BLASTX homology search. The genes and proteins are useful  
CC for preventing, ameliorating or treating medical conditions, e.g. by  
CC protein or gene therapy. The genes are isolated from a range of human  
CC tissues disclosed in the specification. The nucleic acids, proteins,  
CC antibodies and (ant)agonists are useful in the diagnosis, treatment and  
CC prevention of: (a) cancer, e.g. breast and ovarian cancer, and other  
CC cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal  
CC tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's  
CC disease, allergies, autoimmune haemolytic anaemia, autoimmune  
CC thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,  
CC rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders  
CC such as myocardial ischaemias; (d) wound healing; (e) neurological  
CC diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases  
CC such as viral, bacterial, fungal and parasitic infections  
XX  
XX Sequence 1152-AA;  
Query Match 98.3%; Score 5829.5; DB 4; Length 1152;  
Best Local Similarity 97.0%; Pred. No. 0;  
Matches 1117; Conservative 0; Mismatches 0; Indels 35; Gaps 1;  
QY 1 MELPFVTHLFLPLVFLTGLCSPPNLDHHPRLFPQPPAEAFGYSLQHVGGQQRWMLVGA 60  
DB 1 MELPFVTHLFLPLVFLTGLCSPPNLDHHPRLFPQPPAEAFGYSLQHVGGQQRWMLVGA 60  
QY 61 PWDGPGDRRGDVVRCVPGGAHAPCAKGLHGYQLGNSHPVNMHGLMSLLTDDGG 120  
DB 61 PWDGPGDRRGDVVRCVPGGAHAPCAKGLHGYQLGNSHPVNMHGLMSLLTDDGG 120  
QY 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTMYDVIIVLDSNSIYP 180  
DB 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTMYDVIIVLDSNSIYP 180  
QY 181 WSEVQTFLLRLVGLKFLDIDPEIQVGLVOYGSPPHWSLGDPRFKEEVVRAAKNLSRREG 240  
DB 181 WSEVQTFLLRLVGLKFLDIDPEIQVGLVOYGSPPHWSLGDPRFKEEVVRAAKNLSRREG 240  
QY 241 RETTQAQIMVACTEGESQSHGGPPEARLAVVTDGESHDPGEELPAALKACEAGRVTY 300  
DB 241 RETTQAQIMVACTEGESQSHGGPPEARLAVVTDGESHDPGEELPAALKACEAGRVTY 300  
QY 301 GIAVLGHLRQRDPSSFLRIRTIASDPDRFFNFVNTDEAALTDIVDALGDRIFGLGEGS 360  
DB 301 GIAVLGHLRQRDPSSFLRIRTIASDPDRFFNFVNTDEAALTDIVDALGDRIFGLGEGS 360  
QY 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGYDMGSGVLWLEGHRLFPFRMALEDEF 420

DB 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVGYDMGSGVLWLEGHRLFPFRMALEDEF 420  
QY 421 PPALQNHAAAYLGYSVSSMLLRGGRRFLSGAPRFRHGRKVIAFOLKKDGAVRVAQSLOGE 480  
DB 421 PPALQNHAAAYLGYSVSSMLLRGGRRFLSGAPRFRHGRKVIAFOLKKDGAVRVAQSLOGE 480  
QY 481 QIGSYFSGSELCPDITDRDGTDLVLLVAAPMFLGPONKGTGRVYVYLQGOSSLLTLOGLTQ 540  
DB 481 QIGSYFSGSELCPDITDRDGTDLVLLVAAPMFLGPONKGTGRVYVYLQGOSSLLTLOGLTQ 540  
QY 541 PEPPQDARFGFAMGALPDNLQDGFADVAVGAPLEDGCHQGALYLYHGTQSGVRPAPORIA 600  
DB 541 PEPPQDARFGFAMGALPDNLQDGFADVAVGAPLEDGCHQGALYLYHGTQSGVRPAPORIA 600  
QY 601 AASMPHALSIFGRSVDCGRDLDDGDDLVDAVGAQAAILSSRPVHLTPSLEVTPOAIS 660  
DB 601 AASMPHALSIFGRSVDCGRDLDDGDDLVDAVGAQAAILSSRPVHLTPSLEVTPOAIS 660  
QY 661 VWORDCRRRQGEAVCLTAALCFQVTSRTPGRDHOFYMRFTASLDEWTAGARAAFDGSGQ 720  
DB 661 VWORDCRRRQGEAVCLTAALCFQVTSRTPGRDHOFYMRFTASLDEWTAGARAAFDGSGQ 720  
QY 721 RLSPRRRLSVGNVTCBOLHFHVLDTSDYLRPVALTVTFFALDNTTKPGPVLNEGSPISIQ 780  
DB 721 RLSPRRRLSVGNVTCBOLHFHVLDTSDYLRPVALTVTFFALDNTTKPGPVLNEGSPISIQ 780  
QY 781 KLVPFSSKDCGPDNECVTDVLQVNMIDIRGSRKAPFVVRGRRKRVLSTTLENRKENAYNT 840  
DB 781 KLVPFSSKDCGPDNECVTDVLQVNMIDIRGSRKAPFVVRGRRKRVLSTTLENRKENAYNT 840  
QY 841 SLSIIFSRNLHLASLTQPSPIKVECAAPSAHARLCSVGHVPVQTKAKVTFLEFEFSC 900  
DB 841 SLSIIFSRNLHLASLTQPSPIKVECAAPSAHARLCSVGHVPVQTKAKVTFLEFEFSC 900  
QY 901 SLSLSQVFGKLTASSDSLRNGTLQENTAQTSAYIOVEPHLLFSSSESTLHREYVHPYGT 960  
DB 901 SLSLSQVFGKLTASSDSLRNGTLQENTAQTSAYIOVEPHLLFSSSESTLHREYVHPYGT 960  
QY 961 PVGPGPEFKTTLR-----TNNASCIVQNLT 985  
DB 961 PVGPGPEFKTTLRVQNLGCVVSGLIISALLPAVAHGNVFLSLSQVITNNASCIVQNLT 1020  
QY 986 EPPGPPVHPEELQHTNRLNGSNTQCQVVRCHLGOLAKGTEVSVGLLHVHNEFFRAKFK 1045  
DB 1021 EPPGPPVHPEELQHTNRLNGSNTQCQVVRCHLGOLAKGTEVSVGLLHVHNEFFRAKFK 1080  
QY 1046 SLTVVSTFELGTGREGSVLQLEASRWSESLELVVQTRPILISLWILIGSVLGGILLALL 1105  
DB 1081 SLTVVSTFELGTGREGSVLQLEASRWSESLELVVQTRPILISLWILIGSVLGGILLALL 1140  
QY 1106 VFCLWKLGFPAH 1117  
DB 1141 VFCLWKLGFPAH 1152  
RESULT 9  
AAB64658  
ID AAB64658 standard; protein; 1152 AA.  
XX AAB64658;  
AC AAB64658;  
XX 22-MAR-2001 (first entry)  
XX Human secreted protein BLAST search protein SEQ ID NO: 168.  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antiulcer;  
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein.  
XX Homo sapiens.

XX WO200077197-A1.  
 XX 21-DEC-2000.  
 XX 01-JUN-2000; 2000WO-US014934.  
 XX 11-JUN-1999; 99US-0138599P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX (ROSE/) ROSEN C A.  
 XX Rosen CA, Ruben SM, Komatsoulis GA;  
 XX WPI; 2001-032312/04.  
 XX Isolated nucleic acid molecule encoding a human secreted protein is used  
 XX in preventing, treating or ameliorating a medical condition.  
 XX Disclosure; Page 547-551; 558pp; English.  
 XX The invention relates to the isolation of genes AAP32757-F32803 encoding  
 XX the human secreted proteins AB64549-B64594. The sequence is used as a  
 XX query sequence for doing BLASTX searches to identify homologous  
 XX sequences. The genes and proteins are useful for preventing, ameliorating  
 XX or treating medical conditions, e.g. by protein or gene therapy. The  
 XX genes are isolated from a range of human tissues disclosed in the  
 XX specification. The nucleic acids, proteins, antibodies and (ant)agonists  
 XX are useful in the diagnosis, treatment and prevention of: (a) cancer,  
 XX e.g. breast and ovarian cancer, and other cancers of the adrenal gland,  
 XX bone, bone marrow, breast, gastrointestinal tract, liver, lung, or  
 XX urogenital; (b) immune disorders e.g. Addison's disease, allergies,  
 XX autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
 XX Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 XX colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d)  
 XX wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 XX and parasitic infections  
 XX Sequence 1152 AA;  
 XX  
 XX Query Match 97.2%; Score 5763.5; DB 4; Length 1152;  
 XX Best Local Similarity 95.9%; Pred. No. 0;  
 XX Matches 1105; Conservative 3; Mismatches 9; Indels 35; Gaps 1;  
 XX  
 QY 1 MELPFVTHLFLPLVFLTGLCSFNLDEHRLPFGPPPEAFSGYSLQHVGGGRMMLVGA 60  
 DB :|||||  
 QY 1 VELPFVTHLFLPLVFLTGLCSFNLDEHRLPFGPPPEAFSGYSLQHVGGGRMMLVGA 60  
 DB :|||||  
 QY 61 PWDGSGDRGRGVYRCVGGAHNAPCAKGLGDYQLGNSHPAVNMHLGMSLLETDDGG 120  
 DB :|||||  
 QY 61 PWDGSGDRGRGVYRCVGGAHNAPCAKGLGDYQLGNSHPAVNMHLGMSLLETDDGG 120  
 DB :|||||  
 QY 121 FMACAPLWSRACGSSVFSSGTCARVDASFQPGSLAPTAQRCTPTMDVIVLDGNSIYP 180  
 DB :|||||  
 QY 121 FMACAPLWSRACGSSVFSSGTCARVDASFQPGSLAPTAQRCTPTMDVIVLDGNSIYP 180  
 DB :|||||  
 QY 181 WSEVQTFLRLVGLKFLTDPEIQVGLVQYGESPVHWSLGDFTKKEEVVRAAKNLSRREG 240  
 DB :|||||  
 QY 181 WSEVQTFLRLVGLKFLTDPEIQVGLVQYGESPVHWSLGDFTKKEEVVRAAKNLSRREG 240  
 DB :|||||  
 QY 241 RETTQAQIMVACTEGFSQSHGGRPEARLLVVYTDGSHDGBELPAALKACEAGRVTRY 300  
 DB :|||||  
 QY 241 RETTQAQIMVACTEGFSQSHGGRPEARLLVVYTDGSHDGBELPAALKACEAGRVTRY 300  
 DB :|||||  
 QY 301 GIAVLGHLRQRDPSSFLREIRTIASDPDRPFNFVNTDEAALTDIVDALGDRIFGLG 360  
 DB :|||||  
 QY 301 GIAVLGHLRQRDPSSFLREIRTIASDPDRPFNFVNTDEAALTDIVDALGDRIFGLG 360  
 DB :|||||  
 QY 361 HAENESSFGLMSQIGSTHRLKDGILFGMVGAYDWGGSVLMWEGGHLRFPFRMALBDEF 420  
 DB :|||||  
 QY 361 HAENESSFGLMSQIGSTHRLKDGILFGMVGAYDWGGSVLMWEGGHLRFPFRMALBDEF 420  
 DB :|||||

QY 421 PPALQNHAAAYLYGYSVSMLLRGRRRLFLGAPFRHRGKVIAFOLKKDQAVRVAQSLQGE 480  
 DB :|||||  
 QY 421 PPALQNHAAAYLYGYSVSMLLRGRRRLFLGAPFRHRGKVIAFOLKKDQAVRVAQSLQGE 480  
 DB :|||||  
 QY 481 QIGSYFGSELCPDTRDGTDDVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLOQTLO 540  
 DB :|||||  
 QY 481 QIGSYFGSELCPDTRDGTDDVLLVAAPMFLGPQNKETGRVYVYLVGQSSLLTLOQTLO 540  
 DB :|||||  
 QY 541 PEPPQDARFGFAMGALPDNLNODGFADVAVGAPLEDHGHQALYLYHGTQSVRPHPAORIA 600  
 DB :|||||  
 QY 541 PEPPQDARFGFAMGALPDNLNODGFADVAVGAPLEDHGHQALYLYHGTQSVRPHPAORIA 600  
 DB :|||||  
 QY 601 AASMPHALSYFGRSVDRLDGDGDLVDVAVGAQGAAILSSRPVHLTPSLEVTPOAIS 660  
 DB :|||||  
 QY 601 AASMPHALSYFGRSVDRLDGDGDLVDVAVGAQGAAILSSRPVHLTPSLEVTPOAIS 660  
 DB :|||||  
 QY 661 VVORDCRRQGEAVCLTAALCFQVTSRTTCGRWPHQFYMRETASLDEWTAGARAAFDGSGQ 720  
 DB :|||||  
 QY 661 VVORDCRRQGEAVCLTAALCFQVTSRTTCGRWPHQFYMRETASLDEWTAGARAAFDGSGQ 720  
 DB :|||||  
 QY 721 RLSPRRLRLSVGNVTCBOLHFHVLDTSDYLRPVALTVTPALDNTTKPGPVLNBSGSPSIQ 780  
 DB :|||||  
 QY 721 RLSPRRLRLSVGNVTCBOLHFHVLDTSDYLRPVALTVTPALDNTTKPGPVLNBSGSPSIQ 780  
 DB :|||||  
 QY 781 KLVFPSKDCGPDNECVTDLVLQVNMDIRGSRKAPFVVRGGRKVLVSTTLLENKENAYNT 840  
 DB :|||||  
 QY 781 KLVFPSKDCGPDNECVTDLVLQVNMDIRGSRKAPFVVRGGRKVLVSTTLLENKENAYNT 840  
 DB :|||||  
 QY 841 SLSLIIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVCHPVFOTGAKVTFLLPEPFC 900  
 DB :|||||  
 QY 841 SLSLIIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVCHPVFOTGAKVTFLLPEPFC 900  
 DB :|||||  
 QY 901 SSLLSQVFGKLTASSDSLENGTLQENTACTSAIOVEPHLLFSSBSTLHRYEVPYGTLL 960  
 DB :|||||  
 QY 901 SSLLSQVFGKLTASSDSLENGTLQENTACTSAIOVEPHLLFSSBSTLHRYEVPYGTLL 960  
 DB :|||||  
 QY 961 PVGPGPEFKTTLR-----TNNACIVQNLT 985  
 DB :|||||  
 QY 961 PVGPGPEFKTTLR-----TNNACIVQNLT 985  
 DB :|||||  
 QY 986 EPFGPPVHPELOHTNRLNGSNTCCQVVRCHLQOLAKGTESVGLLRLVNEFRRAKPK 1045  
 DB :|||||  
 QY 1021 EPFGPPVHPELOHTNRLNGSNTCCQVVRCHLQOLAKGTESVGLLRLVNEFRRAKPK 1080  
 DB :|||||  
 QY 1046 SLTVVSTFELGTBEGSVLQTEASRWSESLLEVVTQRPILISLWILGSLGGLLLALL 1105  
 DB :|||||  
 QY 1081 SLTVVSTFELGTBEGSVLQTEASRWSESLLEVVTQRPILISLWILGSLGGLLLALL 1140  
 DB :|||||  
 QY 1106 VFCLWKLGPFAH 1117  
 DB :|||||  
 QY 1141 VFCLWKLGPFAH 1152  
 DB :|||||  
 RESULT 10  
 ADR41424  
 ID ADR41424 standard; protein; 1049 AA.  
 XX AC ADR41424;  
 XX 07-OCT-2004 (first entry)  
 XX Human CD-like molecule HA0AD02, SEQ ID NO:223.  
 XX Human; CD-like molecule; cluster of differentiation; diagnosis;  
 XX prevention; immune disorder; immunodeficiency; autoimmune disorder;  
 XX blood-related disorder; haematological disorder; haemostatic disorder;  
 XX thrombolytic disorder; hyperproliferative disorder; cancer; tumour;  
 XX apoptotic disorder; cardiovascular disorder; respiratory disorder;  
 XX angiogenic disorder; neovascularisation; neurological disorder;  
 XX endocrine disorder; reproductive system disorder; infectious disease;  
 XX gastrointestinal disorder; drug screening; tissue regeneration;  
 XX chemotaxis; gene therapy; antibody therapy; drug targeting;  
 XX chromosome mapping; forensic analysis; immunophenotyping; cytostatic;



KW haemostatic; tranquiliser; vulnery; antiinflammatory; nephrotropic;  
 KW cardiac; aniallergic; anti-HIV; antirheumatic; antithyroid;  
 KW antiparietal; immunosuppressive; vasotropic; neurotropic; neuroprotective;  
 KW antithyroid; thyromimetic; gynaecological; virucide; hepatotropic;  
 KW antibacterial; dermatological; chromosome 1q21.

OS Homo sapiens.

XX WO200226930-A2.

XX 04-APR-2002.

XX 25-SEP-2001; 2001WO-US029838.

XX 26-SEP-2000; 2000US-0235484P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Birse CE;

XX WPI; 2002-405050/43.

DR N-PSDB; ADR41248.

XX Novel polynucleotides and polypeptides useful for treating, preventing or  
 PT ameliorating cardiovascular, renal, neurovascular, and autoimmune  
 PT disorders.

XX Claim 11; SEQ ID NO 223; 1243pp; English.

XX The invention relates to 167 novel human CD (cluster of differentiation)-  
 CC like molecules (ADR41388-ADR41563) and to cDNAs encoding them (seqid:11)-

XX Sequence 1049 AA;

Query Match 82.2%; Score 4878.5; DB 5; Length 1049;

Best Local Similarity 91.1%; Pred. No. 0;

Matches 955; Conservative 7; Mismatches 31; Indels 55; Gaps 4;

Qy 124 CAPLWSRA-CGSSVFSFGICARVDASFQPGSLAPTAQRCPTYMDV---VIVLDGNSIY 179

Db 18 CLQFDMCPGCFIPASG-----KPGTHCPTLPNIHGCHCLGWLHSIY 61

Qy 180 PWSVQTFRLRLVGLKFLDPQIQVGLVQGESPVHWSLGDFTKEVYVAAKNLSRRE 239

Db 62 PWSVQTFRLRLVGLKFLDPQIQVGLVQGESPVHWSLGDFTKEVYVAAKNLSRRE 121

Qy 240 GRETKTAQAINVACTEGFSQSHGRPEAARLLVVVTDGESHGDBELPAALKACAGRVTR 299

Db 122 GRETKTAQAINVACTEGFSQSHGRPEAARLLVVVTDGESHGDBELPAALKACAGRVTR 181

Qy 300 YGIAVLGHYLRQRDPSSFLREIRTIASDPDERFFNNVTDEAALTDIVDALGDRIFGLEG 359

Db 182 YGIAVLGHYLRQRDPSSFLREIRTIASDPDERFFNNVTDEAALTDIVDALGDRIFGLEG 241

Qy 360 SHAENESSFGLMSQIGFSTHRLKDGILFGMVGYDWCSSVLWEGGHRLLPFRMALEDE 419

Db 242 SHAENESSFGLMSQIGFSTHRLKDGILFGMVGYDWCSSVLWEGGHRLLPFRMALEDE 301

Qy 420 FPPALQNHAAVLYGSVSSMLRGRRFLSGAPFRHRGKVIATFOLKKGAVRVAQSLQG 479

Db 302 FPPALQNHAAVLYGSVSSMLRGRRFLSGAPFRHRGKVIATFOLKKGAVRVAQSLQG 361

Qy 480 EQIGSYFSGELCPLDTRDGTDLVLLVAAPMFLGPQNKETGRVYVYLVGQOSLLTLQGT 539

Db 362 EQIGSYFSGELCPLDTRDGTDLVLLVAAPMFLGPQNKETGRVYVYLVGQOSLLTLQGT 421

Qy 540 QPEPPQDARFGAMGALPDNLQDGFADVAVGAPLEDGHQALYLYHGTQSGVRPHPAQRI 599

Db 422 QPEPPQDARFGAMGALPDNLQDGFADVAVGAPLEDGHQALYLYHGTQSGVRPHPAQRI 481

Qy 600 AAASMPHALSYFGRSVGRDLDDGDLVDVAVGAGAAIILSSRPVHLTPSLEVTPOAI 659

Db 482 AAASMPHALSYFGRSVGRDLDDGDLVDVAVGAGAAIILSSRPVHLTPSLEVTPOAI 541

Qy 660 SVVQDRCRRRQGEAVCLTAALCFQVTSRTGRMDHQYMRFTASLDWTAGARAFDGS 719  
 Db 542 SVVQDRCRRRQGEAVCLTAALCFQVTSRTGRMDHQYMRFTASLDWTAGARAFDGS 601  
 Qy 720 QRLSPRLRLSVGNVTCQQLHFHVDLSDYLRPVALTPTFALDNTTKPGPVNLNGSPSTSI 779  
 Db 602 QRLSPRLRLSVGNVTCQQLHFHVDLSDYLRPVALTPTFALDNTTKPGPVNLNGSPSTSI 661  
 Qy 780 QKLVPFSKDCGPDNECVTDLVLVQNMDIRSKAPFVVRGRRKVLVSTTLNKRKNAYN 839  
 Db 662 QKLVPFSKDCGPDNECVTDLVLVQNMDIRSKAPFVVRGRRKVLVSTTLNKRKNAYN 721  
 Qy 840 TSLSLIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPFQCAKVTFLLEPEFS 899  
 Db 722 TSLSLIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPFQCAKVTFLLEPEFS 781  
 Qy 900 CSSLLSQVFGKLTASSDSLSLRNGTLQNTAQTSAIIOYEPHLLFSSSTLHRYEVHPYGT 959  
 Db 782 CSSLLSQVFGKLTASSDSLSLRNGTLQNTAQTSAIIOYEPHLLFSSSTLHRYEVHPYGT 841  
 Qy 960 LPVGPGEPEKTTLR-----TNNASCIVQNL 984  
 Db 842 LPVGPGEPEKTTLRVQNLGCVVWSGLIISALLPAVAGHGNVFLSLSQVITNNASCIVQNL 901  
 Qy 985 TEPPGPPVHPPELOHTNRLNGSNTQCVVRCHLQKAGTEVSVGLRLVHNEFRRAKF 1044  
 Db 902 TEPPGPPVHPPELOHTNRLNGSNTQCVVRCHLQKAGTEVSVGLRLVHNEFRRAKF 961  
 Qy 1045 KSLTVSTFELGTEEGSVLQLTASRWSESILVVOVTRPILISLWILIGSVLGGLLLAL 1104  
 Db 962 KSLTVSTFELGTEEGSVLQLTASRWSESILVVOVTRPILISLWILIGSVLGGLLLAL 1021  
 Qy 1105 LVFLCKLGLGFFAHKKIPEEEKREKLEQ 1132  
 Db 1022 LVFLCKLGLGFFAHKKIPEEEKREKLEQ 1049

RESULT 11

AAU14231

ID AAU14231 standard; protein; 1188 AA.

AC AAU14231;

XX 24-OCT-2001 (first entry)

XX Human novel protein #102.

XX Human; novel protein; Antianaemic; osteopathic; antiinflammatory;

XX immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;

XX anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;

XX antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;

XX thrombolytic; immunogen; antibody; gene therapy; neurological disorder;

XX Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;

XX tissue regeneration; immune disorder.

XX Homo sapiens.

XX WO200155437-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US002623.

XX 25-JAN-2000; 2000US-00491404.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-451939/48.

XX N-PSDB; AAS22536.

PT Isolated polypeptides useful for treating anti-inflammatory diseases,  
 XX nervous system disorders, and for regenerating bone and cartilage.

PS Example 4; Page 578-581; 894pp; English.

CC The invention relates to polynucleotides encoding novel human proteins or  
 CC their active domains. The polypeptides, polynucleotides and antibodies  
 CC raised against the polypeptides are used in a method of treatment of a  
 CC mammal and prevention of disorders caused by the aberrant protein  
 CC expression or activity. The polypeptides can be used as molecular weight  
 CC markers, food supplements, and in antibody production. The polypeptides  
 CC are used to identify compounds which bind to the polypeptides.  
 CC Polynucleotides of the invention are used as probes and primers, for  
 CC sequencing, for chromosome or gene mapping, in the production of  
 CC recombinant proteins, and in generating anti-sense DNA or RNA and in gene  
 CC therapy. Polypeptides of the invention can be used to target drugs to a  
 CC tumour, in assays to determine biological activity, to raise  
 CC antibodies/ elicit an immune response, to determine quantitative protein  
 CC levels, as tissue markers, and to isolate receptors or ligands.  
 CC Polypeptides of the invention may also be useful in treating platelet  
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
 CC the proliferation, differentiation and survival of stem cells, as a  
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
 CC fungal infection or from autoimmunity, cancer, allergy, asthma, graft-  
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory  
 CC diseases, nervous system disorders, and infection. The present sequence  
 CC represents a protein of the invention

XX Sequence 1188 AA;

Query Match 39.7%; Score 2355; DB 4; Length 1188;  
 Best Local Similarity 42.3%; Pred. No. 3.38-201; Indels 80; Gaps 16;  
 Matches 506; Conservative 204; Mismatches 406;

QY 1 MELPFVTHLFLVFLVGLCLSPPLNDBHPLPFGPPPEAFEGYVLOHVGQGRWMLVGA 60  
 DB 1 MDLPRLGVVAWALSMLPGFTDTFMDTRKPRVPGSRTAFPGYTVQQHDSGNKMLVGA 60

QY 61 PWDPGDRGRDVRCPVGGHNAAPCAKHGLDGYQLGNSSHPAVNMHGLMSLTTDGGG 120  
 DB 61 PLETNGYQKTGYVKCPV---IHGNTKLNGLRVTLNSVSRKONMRLGLSLATNPKDNS 117

QY 121 FMACAPLWSRACGSSVFSSGTCARVDASFQPGQSLAPTACRPTYMDVIVLDCGNSIYP 180  
 DB 118 FLACPLWSHECGSSYTTGMSRVNSNFRFSKTVAPALQRCQTYMDIVLDCGNSIYP 177

QY 181 WSEVQTLRLRLVGLFIDPEIQVGLVQYGESPVHWSLQDFRTKEEVRAAKNLSRREG 240  
 DB 178 WVEVQHFLINLKKFYICPGQIQGVVQYGEDVVEHFLNDYRSVKDVEAASHIEQGG 237

QY 241 RETTQAQIMVACTEGFSQSHGGRPEARLLVVVTDSGHGDBELPALKAACEAGRVTRY 300  
 DB 238 TETRTAFGIEFARGEAFQK--GGRKGAKKVMIVITDGHSDSPLEKVIQOISERDNTVRY 295

QY 301 GIAVLGHVLRQRDPSSFLRITASDPDRFPFNVTDEAALTDDIVDGLDRIFGLEGS 360  
 DB 296 AVAVLGYNNRGINPEFLNEIKYIADDPDKHFNVTDEAALDKDIVDALGDRIFSLGEGT 355

QY 361 HAENESSFGLMSQIGFSTHRLKDGILFGMWGAYDWGGSVLWLGSHRLPFPRLMALEDEF 420  
 DB 356 N-KNETSFGLMSQTFSSHVVEDGVLLGAVGAYDMNGAVLKETISAGKVIPLRESYLKEF 414

QY 421 PPALQNAHYLYGYSSVSMMLRGRRRLFLSGAPRRHKGKVIAPOLKKGAVRVAAQSLQGE 480  
 DB 415 PEELKNHGAYLYGTVTVSVSSRQGRVYVAGAPRFNHTGKVLFTMHNKSLTIHQAWRGQ 474

QY 481 QIGSYFSGSELCPDTRDGTDDVLLVAAPMLGPNKGTGRVYVYVVGQOQLLTQLQTLQ 540  
 DB 475 QIGSYFSGSEITSDIGDGVTDVLLVAGPMTFN--EGRERKGVYVIEL--KQNRVYVNGTLK 532

QY 541 PEPP-QDARPGFAMGALPDLNQDGFADVAVGAPLEDHQGALYLYHGTQSGVRPHPAQRI 599  
 DB 533 DSHSYQNARFGSSIASVRDLNQDSYNDVVVVGAPLEDNHAGAIYIFHGRGSIILKTPQRI 592

QY 600 AAASMPHALSYFGRSVDRLDGGDLVDVAVGAQGAAILSSRPVHLTPSLVETPQAI 659  
 DB 593 TASELATGLQYFGCSIHQGLDNEGLDIAVAGALNAVILWSRPVQVQINASHFESKI 652

QY 660 SVQVQDCRRRQGBAVCITAAALCFQVTSRTSGRWDHGFYMRFTASLDEWTAGAAAFDGS 719  
 DB 653 NIFRDCRSGRATCLAAFLCTPIFLAHFQTTVIGIRYNATMDERRYPRAHLEGG 712

QY 720 QRSLPRRLRUSGVNVTCEQLHFVLDTSYLRVALVTVPALDNTTKPGVILNKGSTSI 779  
 DB 713 DRFTNRAVLSSQELCERINFVLDTADYVVKPTFSVEYSLEDPDH--GPMLDGDMPTTL 771

QY 780 QKLVPSKDCGPNCEVTDVLQVNDI-----RGRK-----APFVV 817  
 DB 772 RVSVPPMNGCNEDEHCVDPDLVDLARSDLPTAMBYCQRLVKRPAQDCSAYTSLSDTTVFII 831

QY 818 RGGRRKVLVSTTLNKRKENAYNTSLSIIFSRNLHLASLTPQRSPIKVECAAPS--AHAR 875  
 DB 832 ESTQRVAVATLENKENAYSTVLNISQSANLQFASLIQKEDSDGSEICWNERRLQKQ 891

QY 876 LCSVGHVPFOTGAKVTFLLFEFSCSLLSQVFGKLTASSDSLSERNGTLOENTAQTSAI 935  
 DB 892 VCVSVYPPFRAKAKVAPRLDFEFSKSTFLHHLBIELAAGSDSNERDSTKEDNVAPLRFHL 951

QY 936 QYEPHLLPSSSESLHRYEYVHPYCTLP--VCPGPEFKTLRTN----- 975  
 DB 952 KYEADVLFTRSSLSHYEVKPNSSRLERYDYGIPFPFCIFRIQNLGLFPIHMMMKITPI 1011

QY 976 -----NASC-IVQNLTEPPGPPVHPEELOHTNRLNGSNTCCQVVR 1015  
 DB 1012 ATRSGNRLKLRLDPLTDEANTSCNIWGNSTEYPTPVE--EDLRAPOLNHSNDVVSINC 1070

QY 1016 HLQOLAKGTVEVGLRLVHNEFFRAKFKSLTVVSTFELGTBEGSVLQJTEASRWSES 1075  
 DB 1071 NI-RLVNQBEINFNLNLRSLKALKYKSMKIMVNAALQRPSPFIPREEDPSQIV 1129

QY 1076 LEVVQTRPILISLWILIGSVLGLLILALVCLKGLFFAHKKIPEEKREKLE 1131  
 DB 1130 FEISKQEDWQVPIWIVGSLGGLLLALLVLAWLKGLGPPRSAR-----RRREPGLD 1181

RESULT 12  
 AAB50085  
 ID AAB50085 standard; protein; 1188 AA.  
 XX AC AAB50085;  
 XX AC AC  
 DT 19-MAR-2001 (first entry)  
 XX Human A259.  
 DE Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;  
 KW kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;  
 KW rheumatoid arthritis.  
 XX OS Homo sapiens.  
 XX Key Location/Qualifiers  
 FT Domain 1..1141  
 FT /label= Extracellular\_domain  
 FT Peptide 1..22  
 FT /label= Signal\_peptide  
 FT Protein 23..1188  
 FT /label= Mature\_protein  
 FT Domain 39..74  
 FT /label= Integrin\_alpha\_subunit\_repeat\_domain\_#1  
 FT Domain 115..157  
 FT /label= Integrin\_alpha\_subunit\_repeat\_domain\_#2  
 FT Domain 164..345

FT	Domain	/label= I_domain
FT	367..392	
FT	/label= Integrin_alphasubunit_repeat_domain_#3	
FT	421..455	
FT	/label= Integrin_alphasubunit_repeat_domain_#4	
FT	478..516	
FT	/label= Integrin_alphasubunit_repeat_domain_#5	
FT	540..575	
FT	/label= Integrin_alphasubunit_repeat_domain_#6	
FT	602..640	
FT	/label= Integrin_alphasubunit_repeat_domain_#7	
FT	1142..1164	
FT	/label= Transmembrane_domain	
FT	1165..1188	
FT	/label= Cytoplasmic_domain	
XX		
XX	W0200073339-A1.	
XX		
XX	07-DEC-2000.	
XX		
XX	15-MAY-2000; 2000WO-US013262.	
XX		
XX	28-MAY-1999; 99US-00322790.	
PR	27-APR-2000; 2000US-00561363.	
XX		
XX	(MILL-) MILLENNIUM PHARM INC.	
PA		
XX		
PI	Pan Y, Lora JM;	
XX		
XX	WPI; 2001-041142/05.	
DR	N-PSDB; AAC91901, AAC91902.	
XX		
XX	Nucleic acid encoding alpha-integrin subunits, useful for treatment and	
PT	diagnosis of fibrosis, e.g. of the liver.	
PT		
XX		
PS	Claim 8; Fig 1; 164pp; English.	
XX		
CC	The present sequence is human integrin alpha subunit, A259. A259 is	
CC	homologous with the alpha1 and alpha10 integrin subunits and is	
CC	overexpressed in fibrosis. A259 is implicated in regulation of	
CC	proliferation, differentiation and/or function of many different cell	
CC	types. Inhibitors of A259 activity are useful for the treatment of liver	
CC	disease, particularly fibrosis, and also fibrosis in other organs	
CC	(specifically lung and kidney). In addition, A259 can be used for	
CC	treatment and prevention of cancer, osteoporosis, acute myeloid	
CC	leukaemia, HIV infection, and rheumatoid arthritis	
XX		
SQ	Sequence 1188 AA;	
Query Match 39.7%; Score 2355; DB 4; Length 1188;		
Best Local Similarity 42.3%; Pred. No. 3.3e-201;		
Matches 506; Conservative 204; Mismatches 406; Indels 80; Gaps 161		
Qy	1 MELPFVTHLFLPLVFLTGLCSPPFNLDHHPLRFPQPPAEAFGYSVLQHVGGQRMVLVGA 60	
Db	1 MDLPRGLVVAWALSILWPGFTDTFNMDTKPRVIFGSRTAFFGYTVQOHDISGNKWLIVGA 60	
Qy	61 PWDGSGDRGDVTRCPVGGAHNAPCAKGHLDYQLGNSSHPAVNMHLGMSLLETGDGG 120	
Db	61 PLETNGYQKTGDVYKCPV---IHGNCCTKLNLGRVTLTSNVSRKONMRLGLSLATNPDKNS 117	
Qy	121 FMACAPLWSRACGSSVFSSGGICARVDASFOFGSILAPTAQRCPTVMQVIVLDSNSIYP 180	
Db	118 FLACSPWSHECGSSYYITGMCNRSNFRSKTVAPALQRCQTYMDIVIVLDSNSIYP 177	
Qy	181 WSEVQTEFLRLVLGKFLIDPEQIQVLGVQGESPSHWSLGBFRTKEEVVRAAKNLSRREG 240	
Db	178 WVEVQHFLINILKIFYIGPGQIQGVVQYGEDVWHEFLNDYRSVKDVVEAASHIEQSGG 237	
Qy	241 RETYTAQAIWVACTEGFSQSHGGRPPEARLLVWTDGSHDGEELPALAKACEAGRVTRY 300	
Db	238 TETRTAFGIEFARSEAFQK--GGRKGAKVMIVTDGSHDSPDLKVIQGSERDNVTRY 295	



Db 832 ESTQRVAVATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSIQECVNEERRLQKQ 891  
 Qy 876 LCSVCHVPFOGAKVTFLEFEFSCSLLSQVFGKLTASSLSLRNGTLQENTQTSAYI 935  
 Db 892 VCNVSYPPFRKAKVAFRLDFEFSKIFLHLELELAAGSDSNERDSTKEDNVAPLRFHL 951  
 Qy 936 QYEPHLLFSBSSTLHRYVHPVGTLP--VGQPPFKTTLRTN----- 975  
 Db 952 KYEADVLFTRSSLSHYEVKNSSLYERYDVGPPFCIFRQNLGLFIHGMKMTIPI 1011  
 Qy 976 -----NASC-IVQNLTPPPGPPVHPPELQHTNRLNGSNTQCVVRC 1015  
 Db 1012 ATRSGNRLKLRDLFTDEANTSCNIGNSTEVRPTPVE-EDLRRAPQLHNSNDVWSINC 1070  
 Qy 1016 HLGOLAKTEVSGVGLLRVHNEFFRFRKFSKLTVVSTFELTEGSGVQLTEASRWSBSL 1075  
 Db 1071 NI-RLVPNQEFNPHLLGNLWLRSLKALKYKMKIMVNAALQRFHSPPIFREEDPSRQIV 1129  
 Qy 1076 LEVVQTRPILSLWLTIGSVLGLLLALLVFLWKLGFFAHKKIPBEKEEKKLE 1131  
 Db 1130 FEISKQEDWQPIWIIIVGSTLGGLLALLVALLWKLGGFFRSAR-----RRREPGLD 1181

## RESULT 14

AAU14467  
 ID AAU14467 standard; protein; 1188 AA.

XX AAU14467;

AC 24-OCT-2001 (first entry)

XX Human novel protein #338.

XX Human; novel protein; Antianaemic; osteopathic; antiinflammatory;  
 immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;  
 anticonvulsant; antithalric; cerebroprotective; antifungal; antiviral;  
 antibacterial; antiallergic; dermatological; haemostatic; antisthmatic;  
 thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
 Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
 tissue regeneration; immune disorder.

XX Homo sapiens.

XX WO200155437-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US002623.

XX 25-JAN-2000; 2000US-00491404.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-451939/48.

XX N-PSDB; AAS22772.

XX Isolated polypeptides useful for treating anti-inflammatory diseases,

XX nervous system disorders, and for regenerating bone and cartilage.

XX Example 4; Page 828-831; 894pp; English.

XX The invention relates to polynucleotides encoding novel human proteins or  
 their active domains. The polypeptides, polynucleotides and antibodies  
 raised against the polypeptides are used in a method of treatment of a  
 mammal and prevention of disorders caused by the aberrant protein  
 expression or activity. The polypeptides can be used as molecular weight  
 markers, food supplements, and in antibody production. The polypeptides  
 are used to identify compounds which bind to the polypeptides.

XX Polynucleotides of the invention are used as probes and primers, for  
 sequencing, for chromosome or gene mapping, in the production of  
 recombinant proteins, and in generating anti-sense DNA or RNA and in gene

CC therapy. Polypeptides of the invention can be used to target drugs to a  
 tumour, in assays to determine biological activity, to raise  
 antibodies/elicit an immune response, to determine quantitative protein  
 levels, as tissue markers, and to isolate receptors or ligands.  
 CC Polypeptides of the invention may also be useful in treating platelet  
 disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
 ligament and/or nerve tissue, wound healing, treating burns, promoting  
 the proliferation, differentiation and survival of stem cells, as a  
 contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
 sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
 fungal infection or from autoimmunity, cancer, allergy, asthma, graft-  
 CC versus-host disease, eczema, haemophilia, thrombosis, anti-inflammatory  
 diseases, nervous system disorders, and infection. The present sequence  
 CC represents a protein of the invention  
 XX

SQ Sequence 1188 AA;

Query Match 39.7%; Score 2354; DB 4; Length 1188;

Best Local Similarity 42.3%; Pred. No. 4.le-201;

Matches 506; Conservative 204; Mismatches 406; Indels 80; Gaps 16;

Qy 1 MELPFVTHLFLPLVFLTGLCSFNLDEHPRLPFGPEAEFGYSLQHVGGQRMVLVGA 60  
 Db 1 MDLPRGLVVAWALSFWGFTDTFNMDTRKPRVIFGSRATFEGYTVQQHDSGNKWLVLVGA 60  
 Qy 61 PWDGSGDRRGDYYRCPVGGAHNAPCAKHGLDYLQNGNSHPAVNMHLMGMLLETDDGG 120  
 Db 61 PLETNGYQKTGDVYKCPV---IHGNCITKLNLGRVTLNVSERKONMRGLSLATNPKDNS 117  
 Qy 121 FMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVVILDCGNSLYP 180  
 Db 118 FLACSPWSHECGSSYYTGMCSRNFRSKTVAPALORCQTYMDIVIVLDGNSLYP 177  
 Qy 181 WSEVQTELRRLVGLKFLIDPEIQVGLVQYGESPVHMSLGDGFRTEKEEVRAAKNLSRREG 240  
 Db 178 WVEVQHEFLINLLKFIYGPQIQVGVVQYGEDVVEHFLNDYRSVKDVVEAASHIEQREG 237  
 Qy 241 RETKTAQIAIVACTEGFSQSHGGRPEARLLVVVTDSHGDGEBELPAALKAACEAGRVTY 300  
 Db 238 TETRTAFGIEFARSEAFQK--GGRKGAQVMIVITDGHSDSPDLKVIQOOSRDNVTRY 295  
 Qy 301 GIAVLGHYLRQRDPSSFLREIRTIASDPDERPFNVTDEAALTDIVDALGDRIFGLEGS 360  
 Db 296 AVAVLYGNRRGINPETFLNEIKYIASDPDDKGFNVTDEAALXDIVDALGDRIFSLEGT 355  
 Qy 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVAGYDWGSLVLEGGHRLPPRMALEDEF 420  
 Db 356 N-KNETSFGLMSQTFSSHHVEDGVLGAVGAYDWNGAVLKETSAGKVIPLRESYLKEF 414  
 Qy 421 PPALQNHAAAYLGYSVSMLLRGGRRFLSGAPRFRHGRKVIQFQKKDGAVRVAQSLQGE 480  
 Db 415 PEELKNHGAYLGYTVTSVSSRQGRVVVAGAPRFNHTGKVLFTMNNRSLTIHQAMRGQ 474  
 Qy 481 QIGSYFSGELCPDLTDRDGTDLVLLVAAPFLGPONKETGRVYVYLVGQQSLLTQGLTQ 540  
 Db 475 QIGSYFSGEITSVDIDGDTVDLLVAGAPMYFN--EGREGKVVYVEL--RQNRVYVNGTLK 532  
 Qy 541 PEPP-QDARFGFAMGALPDNLQDGFADVAVGAPLEDHOGALVLYHGTQSGVRPHPAORI 599  
 Db 533 DSHSYQNRATFGSSIASVRDLNQDSYNDVVGAPLEDNHAGAIYIFHGFSGILKTPKORI 592  
 Qy 600 AAASMPHALSYFGRSVDGRDLDDGLVDVAVAGQGAAILSSRPFIHLTPSLEVTPQAI 659  
 Db 593 TASELATGLQYFGCSIHGQDLNEDGLIDLAVGALGNAVILWSRPVQINASLHFPESKI 652  
 Qy 660 SVVORDCRRGQEAIVCLTALCFQVTSRTSGRWDHQYMRPTASLDEWAGARAAFDGSG 719  
 Db 653 NIFHRDCKRSGRDATCLAAFLCFTPIFLAPHFQTTTGVIRYNATMDERRYPRAHLDEGG 712  
 Qy 720 QRLSPRLRLSVGNVTCEQLHFEHVLDTSDYLRLPVALTVTTFALDNTTKPGPVNLSGSPSI 779  
 Db 713 DRFTNRVALLSSGQELCERINFHVLDTADYVYKVTTFSEYSLSDPDH-GPMLDDGWPTTL 771



Db 832 ESTQRVAVATLENRGENAYSTVLNISQSANLOFASLIQKEDSDGSIECVNEERRLOKQ 891  
Qy 876 LCSVGHVPVOTGAKVTFLLBEPFSCSSLLSQVFGKLTASSDLSLERNCTLOENTAQTSAYI 935  
Db 892 VCNVSYPPFFRAKAKVAFRLDFEFKSIFLHHLEIETELAAGSDSNERDSTKEDNVAPLRFHL 951  
Qy 936 QYEPHLLFSSSESTLHRYEVHPYGTLP--VGPGPEFKTTLRTN----- 975  
Db 952 KYEVDVLFTRSSLSLHYEVKPNSSLERYDYGIGPPFCIFRIQNLGLPFIHGMMMKITIPI 1011  
Qy 976 -----NASC-IVQNLTEPPGPPVHPPEELOHTNRLNGSNTQCQVRC 1015  
Db 1012 ATRSGNRLKLRLDFLTDEANTSCNIGNSTEYRPTVE-EDLRAPQLNHSNDVVSINC 1070  
Qy 1016 HLGQLAKGTEVSVGLLRLVHNEFFRAKFKSLTVVSTFELGTEGVSVLQLTEASRWSESL 1075  
Db 1071 NI-RLVPNQEIHFLLGNLWLSLKALKYKSMKIMVNAALQRFHSPFIFREEPDSRQIV 1129  
Qy 1076 LEVVQTRPILISLWILIGSVLGGLLALLVFCILWKLGFPAHKKIPPEEKREEKLE 1131  
Db 1130 FEISKQEDWQVPIWIVGSTLGGLLALLVLAALWKLGFPSAR-----RRREPGLD 1181

Search completed: April 6, 2005, 12:22:35  
Job time : 149.162 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:17:21 ; Search time 141.389 Seconds

(without alignments)

2658.060 Million cell updates/sec

Title: US-09-647-544-4

Perfect score: 5932

Sequence: 1 MELPFVTHLFLPLVLTGLC.....GFAHKKIPREKREKLEQ 1132

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1418010 seqs, 331997259 residues

Total number of hits satisfying chosen parameters: 1418010

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubaa/US09\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubaa/US10\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubaa/US10D\_PUBCOMB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubaa/US11\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubaa/US11\_NEW\_PUB.pep:\*
- 19: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep:\*
- 20: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5890.5	99.3	1167	16	US-10-741-601-531
2	5890.5	99.3	1177	16	US-10-741-601-532
3	2355	39.7	1188	15	US-10-291-265-338
4	2354	39.7	1188	15	US-10-291-265-810
5	2344.5	39.5	1189	10	US-09-984-130-35
6	2344.5	39.5	1189	10	US-09-836-353A-35
7	2344.5	39.5	1189	15	US-10-262-839-4
8	2194	37.0	1034	10	US-09-984-130-43
9	2194	37.0	1034	10	US-09-836-353A-43
10	2140	36.1	1120	15	US-10-262-839-6
11	1897.5	31.8	1179	10	US-09-918-715-250
12	1884	31.8	1180	10	US-09-918-715-307
13	1881.5	31.7	1151	10	US-09-984-130-103
					Sequence 531, App
					Sequence 532, App
					Sequence 338, App
					Sequence 810, App
					Sequence 35, Appl
					Sequence 35, Appl
					Sequence 4, Appl
					Sequence 43, Appl
					Sequence 43, Appl
					Sequence 6, Appl
					Sequence 250, App
					Sequence 107, App
					Sequence 103, App

14	1881.5	31.7	1151	10	US-09-836-353A-103	Sequence 103, App
15	1874.5	31.6	1177	17	US-10-741-600-1161	Sequence 1161, App
16	1733.5	29.2	1181	14	US-10-160-354-2	Sequence 2, Appl
17	1733.5	29.2	1181	15	US-10-295-027-1286	Sequence 1286, App
18	1733.5	29.2	1181	15	US-10-211-462-187	Sequence 187, App
19	1727.5	29.1	707	9	US-09-764-870-313	Sequence 313, App
20	1727.5	29.1	707	14	US-10-125-540-313	Sequence 313, App
21	1719.5	29.0	1147	15	US-10-336-603A-42	Sequence 42, Appl
22	1718	29.0	1148	16	US-10-872-198-147	Sequence 147, App
23	1126	19.0	1161	9	US-09-350-259-2	Sequence 2, Appl
24	1124	18.9	1161	10	US-09-891-943-2	Sequence 2, Appl
25	1124	18.9	1161	9	US-09-350-259-55	Sequence 55, Appl
26	1124	18.9	1161	10	US-09-891-943-55	Sequence 55, Appl
27	1120.5	18.9	1161	9	US-09-350-259-99	Sequence 99, Appl
28	1120.5	18.9	1161	10	US-09-891-943-99	Sequence 99, Appl
29	1116.5	18.8	1170	17	US-10-741-600-1088	Sequence 1088, App
30	1115.5	18.8	1170	9	US-09-945-265-2	Sequence 1, Appl
31	1114.5	18.8	1170	15	US-10-261-164-1	Sequence 1, Appl
32	1114.5	18.8	1161	9	US-09-350-259-53	Sequence 53, Appl
33	1114.5	18.8	1161	10	US-09-891-943-53	Sequence 53, Appl
34	1111	18.7	1145	16	US-10-872-198-130	Sequence 130, App
35	1110	18.7	1223	16	US-10-408-765A-295	Sequence 295, App
36	1110	18.7	1223	17	US-10-741-600-1086	Sequence 1086, App
37	1108.5	18.7	1155	9	US-09-350-259-46	Sequence 46, Appl
38	1108.5	18.7	1155	10	US-09-891-943-46	Sequence 46, Appl
39	1108	18.7	1151	9	US-09-350-259-37	Sequence 37, Appl
40	1108	18.7	1151	10	US-09-891-943-37	Sequence 37, Appl
41	1091	18.4	1153	9	US-09-350-259-3	Sequence 3, Appl
42	1091	18.4	1153	10	US-09-902-481A-1	Sequence 1, Appl
43	1091	18.4	1153	10	US-09-891-943-3	Sequence 3, Appl
44	1091	18.4	1153	14	US-10-144-259-30	Sequence 30, Appl
45	1091	18.4	1153	14	US-10-207-655-176	Sequence 176, App

#### ALIGNMENTS

#### RESULT 1

US-10-741-601-531  
; Sequence 531, Application US/10741601  
; Publication No. US20040166519A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; FILE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01500  
; CURRENT APPLICATION NUMBER: US/10/741.601  
; CURRENT FILING DATE: 2003-12-22  
; NUMBER OF SEQ ID NOS: 26415  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 531  
; LENGTH: 1167  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-741-601-531

Query Match 99.3%; Score 5890.5; DB 16; Length 1167;  
Best Local Similarity 96.7%; Pred. No. 0;  
Matches 1129; Conservative 2; Mismatches 1; Indels 35; Gaps 1;

Qy	1	MELPFVTHLFLPLVLTGLCSPNLDHHPRLPGPPEAEFGYSLQHVGGQRMVLGA	60
Db	1	MELPFVTHLFLPLVLTGLCSPNLDHHPRLPGPPEAEFGYSLQHVGGQRMVLGA	60
Qy	61	PWDGPGSDRRGDVYRCVPVGAHNAKCAKGLDGYQLGNSHPAVNMHGLMSLLTDDGGG	120
Db	61	PWDGPGSDRRGDVYRCVPVGAHNAKCAKGLDGYQLGNSHPAVNMHGLMSLLTDDGGG	120
Qy	121	FMACAPLWSRACGSSVFSSGICARVDASFPQGSAPTAQRCPYMDVIVLDGNSNIYP	180
Db	121	FMACAPLWSRACGSSVFSSGICARVDASFPQGSAPTAQRCPYMDVIVLDGNSNIYP	180
Qy	181	WSEVQTFRLRLVKGKLFIDPQIQVGLVQYGSPVHWSLGDFTKEEVRAAKNLSREG	240

```
Db 181 WSEVQTLRLVLGKLFIDPEQIQVGLVQYGESPVHWSLGDFTKEVRAAKNLSRREG 240
Qy 241 RETKTAQAINVACTEGFSQSHGGRPEAARLLVVVTDGESHGDBELPAALKACAGRVTRY 300
Db 241 RETKTAQAINVACTEGFSQSHGGRPEAARLLVVVTDGESHGDBELPAALKACAGRVTRY 300
Qy 301 GIAVLGHLRQRDPSSFLREIRTIASDPDERFFNVNTDEAALTDIVDALGDRIFGLEGS 360
Db 301 GIAVLGHLRQRDPSSFLREIRTIASDPDERFFNVNTDEAALTDIVDALGDRIFGLEGS 360
Qy 361 HAENESSFGLMSQIGFSTHRLKXGILFGMVAGYDMGGSVLMLEGGHRLPPRMALEDEF 420
Db 361 HAENESSFGLMSQIGFSTHRLKXGILFGMVAGYDMGGSVLMLEGGHRLPPRMALEDEF 420
Qy 421 PPALQNHAAVLYGYSVSMLLRGGRRLFLSGAPFRHRGKVIAPQLKXDGAVRVAQSLQGE 480
Db 421 PPALQNHAAVLYGYSVSMLLRGGRRLFLSGAPFRHRGKVIAPQLKXDGAVRVAQSLQGE 480
Qy 481 QIGSYFGSELCPDTRDGTDDVLLVAAPMFLGPQNKETGRVYVYLVGQOSSLTLQGTLLQ 540
Db 481 QIGSYFGSELCPDTRDGTDDVLLVAAPMFLGPQNKETGRVYVYLVGQOSSLTLQGTLLQ 540
Qy 541 PEPPQDARFGFAMGALPDLNQDGFADVAVGAPLEDHGQGALYLYHGTSQGVRRPHPAQR 600
Db 541 PEPPQDARFGFAMGALPDLNQDGFADVAVGAPLEDHGQGALYLYHGTSQGVRRPHPAQR 600
Qy 601 AASMPHALSYFGRSVDRGLDLDGDDLDVAVAGQAAILLSRPIVHLTPSLEVTPOAIS 660
Db 601 AASMPHALSYFGRSVDRGLDLDGDDLDVAVAGQAAILLSRPIVHLTPSLEVTPOAIS 660
Qy 661 VVQDCCRROGAEVCLTAALCFQVTSRTPGRWHDQFYMRFTASIDEWNTAGARAAFDGSGQ 720
Db 661 VVQDCCRROGAEVCLTAALCFQVTSRTPGRWHDQFYMRFTASIDEWNTAGARAAFDGSGQ 720
Qy 721 RLSPRRLRLSGVNTCEQLHFHVLDTSDYLRLPVALTVTFALDNTTKPGPVLINEGSPS 780
Db 721 RLSPRRLRLSGVNTCEQLHFHVLDTSDYLRLPVALTVTFALDNTTKPGPVLINEGSPS 780
Qy 781 KLVPFSDKDCPDNECVTDVLQVNMDIRGSRKAPFVVRGRRKVLVSTTLLENKENAYNT 840
Db 781 KLVPFSDKDCPDNECVTDVLQVNMDIRGSRKAPFVVRGRRKVLVSTTLLENKENAYNT 840
Qy 841 SLSLIFSRNLHLASLTTPQRESPIKVECAAPSAHARLCSVGHVPVQTCAGKVTFLLEFFSC 900
Db 841 SLSLIFSRNLHLASLTTPQRESPIKVECAAPSAHARLCSVGHVPVQTCAGKVTFLLEFFSC 900
Qy 901 SLLLSQVFGKLTASSDSLERNGTLQENTAOATSAYIQIEPHLLPSSSESTLHRYEHPYGT 960
Db 901 SLLLSQVFGKLTASSDSLERNGTLQENTAOATSAYIQIEPHLLPSSSESTLHRYEHPYGT 960
Qy 961 PVGPGPFKTTLR-----TNNASCIQVNL 985
Db 961 PVGPGPFKTTLRVQNLGCVYVSLIISALLPAVAHGNYFLSLSQVITNNASCIQVNL 1020
Qy 986 EPPGPPVHPBELQHTNRLNGSNTCCQVVRCHLGQAKGTESVGLLRLVHNEFFRRAKFK 1045
Db 1021 EPPGPPVHPBELQHTNRLNGSNTCCQVVRCHLGQAKGTESVGLLRLVHNEFFRRAKFK 1080
Qy 1046 SLTVVSTFELGTEGSGVLQTEASRWSESLEVVQTRPILISMLIGSVLGGLLALL 1105
Db 1081 SLTVVSTFELGTEGSGVLQTEASRWSESLEVVQTRPILISMLIGSVLGGLLALL 1140
Qy 1106 VFCLWKLGFPAHKIPREEKREKLEQ 1132
Db 1141 VFCLWKLGFPAHKIPREEKREKLEQ 1167
```

## RESULT 2

US-10-741-601-532

; Sequence 532, Application US/10741601

; Publication No. US20040166519A1

; GENERAL INFORMATION:

```
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001500
; CURRENT APPLICATION NUMBER: US/10/741,601
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 532
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-741-601-532
```

Query Match 99.3%; Score 5890.5; DB 16; Length 1177;

Best Local Similarity 96.7%; Pred. No. 0;

Matches 1129; Conservative 2; Mismatches 1; Indels 35; Gaps 1;

```
Qy 1 MELPFFVTHLFLPLVFLTGLCSPFNLDHHPRLPPGPEAFYGVSLQHVGGQRMVLVGA 60
Db 1 MELPFFVTHLFLPLVFLTGLCSPFNLDHHPRLPPGPEAFYGVSLQHVGGQRMVLVGA 60
Qy 61 PWDGSGDRRGDYVRCVPGGAHNAPCAKGLHLDYQLGNSHHPAVNMHLGMSLLETDDGG 120
Db 61 PWDGSGDRRGDYVRCVPGGAHNAPCAKGLHLDYQLGNSHHPAVNMHLGMSLLETDDGG 120
Qy 121 FMACAPLMSRACGSSVFSSGICARVDASFPQGS LAPTAQRCPTYMDVTVLVDGNSIYP 180
Db 121 FMACAPLMSRACGSSVFSSGICARVDASFPQGS LAPTAQRCPTYMDVTVLVDGNSIYP 180
Qy 181 WSEVQTLRLVLGKLFIDPEQIQVGLVQYGESPVHWSLGDFTKEVRAAKNLSRREG 240
Db 181 WSEVQTLRLVLGKLFIDPEQIQVGLVQYGESPVHWSLGDFTKEVRAAKNLSRREG 240
Qy 241 RETKTAQAINVACTEGFSQSHGGRPEAARLLVVVTDGESHGDBELPAALKACAGRVTRY 300
Db 241 RETKTAQAINVACTEGFSQSHGGRPEAARLLVVVTDGESHGDBELPAALKACAGRVTRY 300
Qy 301 GIAVLGHLRQRDPSSFLREIRTIASDPDERFFNVNTDEAALTDIVDALGDRIFGLEGS 360
Db 301 GIAVLGHLRQRDPSSFLREIRTIASDPDERFFNVNTDEAALTDIVDALGDRIFGLEGS 360
Qy 361 HAENESSFGLMSQIGFSTHRLKXGILFGMVAGYDMGGSVLMLEGGHRLPPRMALEDEF 420
Db 361 HAENESSFGLMSQIGFSTHRLKXGILFGMVAGYDMGGSVLMLEGGHRLPPRMALEDEF 420
Qy 421 PPALQNHAAVLYGYSVSMLLRGGRRLFLSGAPFRHRGKVIAPQLKXDGAVRVAQSLQGE 480
Db 421 PPALQNHAAVLYGYSVSMLLRGGRRLFLSGAPFRHRGKVIAPQLKXDGAVRVAQSLQGE 480
Qy 481 QIGSYFGSELCPDTRDGTDDVLLVAAPMFLGPQNKETGRVYVYLVGQOSSLTLQGTLLQ 540
Db 481 QIGSYFGSELCPDTRDGTDDVLLVAAPMFLGPQNKETGRVYVYLVGQOSSLTLQGTLLQ 540
Qy 541 PEPPQDARFGFAMGALPDLNQDGFADVAVGAPLEDHGQGALYLYHGTSQGVRRPHPAQR 600
Db 541 PEPPQDARFGFAMGALPDLNQDGFADVAVGAPLEDHGQGALYLYHGTSQGVRRPHPAQR 600
Qy 601 AASMPHALSYFGRSVDRGLDLDGDDLDVAVAGQAAILLSRPIVHLTPSLEVTPOAIS 660
Db 601 AASMPHALSYFGRSVDRGLDLDGDDLDVAVAGQAAILLSRPIVHLTPSLEVTPOAIS 660
Qy 661 VVQDCCRROGAEVCLTAALCFQVTSRTPGRWHDQFYMRFTASIDEWNTAGARAAFDGSGQ 720
Db 661 VVQDCCRROGAEVCLTAALCFQVTSRTPGRWHDQFYMRFTASIDEWNTAGARAAFDGSGQ 720
Qy 721 RLSPRRLRLSGVNTCEQLHFHVLDTSDYLRLPVALTVTFALDNTTKPGPVLINEGSPS 780
Db 721 RLSPRRLRLSGVNTCEQLHFHVLDTSDYLRLPVALTVTFALDNTTKPGPVLINEGSPS 780
Qy 781 KLVPFSDKDCPDNECVTDVLQVNMDIRGSRKAPFVVRGRRKVLVSTTLLENKENAYNT 840
Db 781 KLVPFSDKDCPDNECVTDVLQVNMDIRGSRKAPFVVRGRRKVLVSTTLLENKENAYNT 840
```

```
Qy 841 SLSIIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPVQTGAKVTFLLFEPFSC 900
Db 841 SLSIIFSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVPVQTGAKVTFLLFEPFSC 900
Qy 901 SLSLSQVFGKLTASSDSLSRNGTLQENTAOQTSAYIQVEPHLLFSSSTLHRYEVPYGTLL 960
Db 901 SLSLSQVFGKLTASSDSLSRNGTLQENTAOQTSAYIQVEPHLLFSSSTLHRYEVPYGTLL 960
Qy 961 PVGPGPFKFTLR-----TNNASCIVQNTL 985
Db 961 PVGPGPFKFTLRVQNLGCVYVSGLIISALLPAVAHGGNYFLSLSQVITNNASCIVQNTL 1020
Qy 986 EPPGPPVHPELOHTNRLNGSNTCCOVVRCHLQOLAKGTEVSGLLRLVHNEPFRRAKFK 1045
Db 1021 EPPGPPVHPELOHTNRLNGSNTCCOVVRCHLQOLAKGTEVSGLLRLVHNEPFRRAKFK 1080
Qy 1046 SLTVVSTFELGTEGSGVLQTEASRWSESLEVVQTRPILISLWLGSLGGLLLALL 1105
Db 1081 SLTVVSTFELGTEGSGVLQTEASRWSESLEVVQTRPILISLWLGSLGGLLLALL 1140
Qy 1106 VFCLWKLGFFAHKKIPBEKEEKLQ 1132
Db 1141 VFCLWKLGFFAHKKIPBEKEEKLQ 1167

RESULT 3
US-10-291-265-338
; Sequence 338, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 338
; LENGTH: 1188
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-338

Query Match 39.7%; Score 2355; DB 15; Length 1188;
Best Local Similarity 42.3%; Pred. No. 2.6e-199;
Matches 506; Conservative 204; Mismatches 406; Indels 80; Gaps 16;

Qy 1 MELPFVTHLFLVFLTGLCSPPFNLDHHPRLFCGPPPEAFSGYVLQHVGGQRMWLVGA 60
Db 1 MDLPRGLVVAWALSILWPGFTTFNMDTRKPRVIFGSRTPAFGYTVQHDISGNKWLWVGA 60
Qy 61 PWDGSPGRRGDVYRCPVGGAAHAPCAKHGLGDYQLGNSSHPAVNMHIGMSLLETGDGG 120
Db 61 PLETNGYQKTGDVYKCPV---IHGNTCKNLGRVTLNSVSRKDNMRGLSLATNPXDNS 117
Qy 121 FMACAPLWSRACGSSVFSSGICARVDASFQGGSLAPTAQRCPTMDVWVILDGNSIYP 180
Db 118 FLACSPLMSHCGSSYTTTQCSRVNSNFRSKTVAPALQRCQTYMDIVILDGNSIYP 177
Qy 181 WSEVQTFLLRLVGLKFLDIPDQIQVGLVOYGSFVHWSLGDGFRTKEEVVRAAKNLSREG 240
Db 178 WVEVQHFLINILKKFYICPGQIQGVVQYGEDVHVEHFLNDYRSVKDVVEASHIEQRGG 237
```

```
Qy 241 RETKTAQIMVACTEGFSQSHGGRPEARLLVVVTDCGSHDGEELPAALKACEAGRVTY 300
Db 238 TETRTAPGIFARSEAFQK--GGRKGAKKVMIVITDGHSDSPLEKVIQOESRDNVTY 295
Qy 301 GIAVLGHYLRQRDPDFSLREIRTIASDPDPERFFNFVTDEAALTDIVDALGDRIFGLEGS 360
Db 296 AVAVLGYNRRGINPETFLEIKYIASDDPKHFFNFVTDEAALKDIVDALGDRIFSLEGT 355
Qy 361 HAENESSFGLMSQIGFSTHRLKDGILFGWVGAYDWCWSVLWLEGGHRLPPRMALEDEF 420
Db 356 N-KNETSFGLMSQTFSSHVVEDVLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEF 414
Qy 421 PPALONHAAVLYGSVSSMLLRGGRRLFLSGAPRRHRGKVIAPOLKKGAVRVAQSLQGE 480
Db 415 PEELKNHAGLYTIVTSVSSRQGRVYVAGAPRNFHTGKVLFTMHNRSITIIQAMRGQ 474
Qy 481 QIGSYFGSELCPDLTDRDGTDLVLLVAAPMFLGPNQKQKTRVYVYLVGQQSLLTQGTLLQ 540
Db 475 QIGSYFGSEITSVDIDGQDVTDLVVGAPMYFN--EGRERGKVVYVEL--RQNFVYNGTLK 532
Qy 541 PEP--QARFGFANGALPDLNODGFADVAGAPLEDHGHQALYLYHGTQSGVRPHPAQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNQDSYNDVWVVGAPLEDNHAGAIYIFHGFPGSILKTPKQRI 592
Qy 600 AAASMPHALSYFGSVDGRDLDDGDLVDVAVAGQGAAILLSRPVHLTPSLEVTPOAI 659
Db 593 TASLATGLQYFGCSIHGQDLNEDGLIDLAVAGNAVLWSPVQVAINSLHFEPSKI 652
Qy 660 SVVORDCRRRGQEAVALCTAALCFQVTSRTPGMDHQFMYMRTASLDEWTAGARAAFDGSG 719
Db 653 NIFHRDCKRGRDATCLAAFLCTFPIFLAHFQITTVGIRYNATMDERRYTTPRAHLDGG 712
Qy 720 QRLSPRLRLSVGNVTCQELHFLVLDTSYLRPVALVTTFALDNTTKPGVNLGSGSPSI 779
Db 713 DRFTNRAVLLSSGQELCERINFHVLDTADYVVKPVTFSVEYSLEDPDH--GPMLDDGWPTL 771
Qy 780 QKLVPFKDCGPDNECVTDVLQVNMDI-----RGSRK-----APVW 817
Db 772 RVSVFFWNGCNEDEHCVDPDLVDARSDLPTAMEYQORVLRKPAQDCSAYTSLFDTTVFII 831
Qy 818 RGRKRVLVSTTLNKRKENAYNTSLIIFSRNLHLASLTPORESPIKVECAAPS--AHAR 875
Db 832 ESTRQVAVATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSIETCVNEERRLOKQ 891
Qy 876 LCSVGHVPVOTGAKVTFLLFEPFSCSLLSQVFKLTASSDSLSRNGTLQENTAOQTSAYI 935
Db 892 VCNVSYPPFRAKAKAVAFRLDPEFSKIFLHLHLELAAGSDSNERDSTKEDNAPLRFHL 951
Qy 936 QYEPHLLFSSSESTLHRYEVPYGTLP--VGPGPEFKTTLRN----- 975
Db 952 KYEADVLTFRSSLSHYEVKPNSSLSRYDGGIGGPPFSCIFRIQNLGLPFIHGMMKTIPI 1011
Qy 976 -----NASC-IVQNLTEPPGPPVHPELOHTNRLNGSNTCCOVVRC 1015
Db 1012 ATRSGNRLKLRLDPLTDEANTSCNIWGNSTYRTPVE--EDLRAPQLNHSNDVVSINC 1070
Qy 1016 HLQOLAKGTEVSGLLRLVHNEPFRRAKFSLTAVSTFELGTEGSGVLQTEASRWSSSL 1075
Db 1071 NI-RLVNEQIEINFLHGLNWLRSUKALKYKSMKIMVNAALQORQPHSPPIFREEDPSRQIV 1129
Qy 1076 LEVVQTRPILISLWLGSLGGLLLALLVFLWKLGFFAHKKIPBEKEEKLQ 1131
Db 1130 FEISKQBDWQVPIIIVGSLTGGLLLALLVFLWKLGFPRARS-----RRREPGLD 1181

RESULT 4
US-10-291-265-810
; Sequence 810, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
```



```
Db 118 FLACPLWSEHCSSYYTTCMCSRVSNNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYP 177
Qy 181 WSEVQTFRLRLVGLKFLDIPQIOVGLVQYGESPVHWSLGDFTKBEVRAAKNLSRREG 240
Db 178 WVEVQHFLINILKFFYIGPGQIQVGVVQYGEDVHVEFLNDYRSVKDVEAASHIEQRGG 237
Qy 241 RETKTAQIIVACTEGFSQSHGGRPEARLLVVVTDGSHDGEELPAALKACEAGRTRY 300
Db 238 TETRTAFGIFARSEAFQK--GGRKGAKKVMIVITDGSHDSDPLEKVIQOOSERDNVTRY 295
Qy 301 GIAVLGHLRRQDRPSSFLREIRTIASDPDERFFNVTDGSHDSDPLEKVIQOOSERDNVTRY 295
Db 296 AVAVLGYNRRGINPETFLNEIKYIASDDPKHFFNVTDGSHDSDPLEKVIQOOSERDNVTRY 295
Qy 361 HAENESSFGLMSQIGSFTHRLKDGILFGMVAGYDWCSSVLMWEGGHLRFPFRMALEDEF 420
Db 356 N-KNETSFGLMSQIGSFTHRLKDGILFGMVAGYDWCSSVLMWEGGHLRFPFRMALEDEF 420
Qy 421 PPALONHAAYLVGSVSMLLRGGRRLFLSGAPRFRHKGKVIATOLKKGAVRVAOSLOGE 480
Db 415 PEEKLNHGAYLVGTVTSVSSRQGRVYVAGAPRFRHKGKVIATOLKKGAVRVAOSLOGE 480
Qy 481 QIGSYFGSEITSDVIDGQVTDVLLVAGAPMYFN--EGRERKVVYVEL--RQNRFYNGTLK 532
Db 475 QIGSYFGSEITSDVIDGQVTDVLLVAGAPMYFN--EGRERKVVYVEL--RQNRFYNGTLK 532
Qy 541 PEPP-QDARFGFAMGALPDNLQDGFADVAVGAPLEDGHOGALYLYHGQTSGVRPHPAQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNQDSYNDVAVGAPLEDNHAGAIYIFHGRFSILKTPKQRI 592
Qy 600 AASMPHALSYFGSRVDRGLDGLDGLVAVAGAGAAIILSSRPVHLTPSLEVTQAI 659
Db 593 TASELATGLQYFGCSIHGQDLNEDGLIDLAVGALNAVILWSRPVQINASHLFEPSKI 652
Qy 660 SVVQDCCRREGOEAVCLTAALCFQVTSRTPCRWHDHOFYMRFTASLDGWTAGARAFDGG 719
Db 653 NIFHRDCKSRGRDATCLAAFLCTPIFLAPHFQTTVGIIRYNATMDERRYPTRAHLDEGG 712

Db 118 FLACPLWSEHCSSYYTTCMCSRVSNNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYP 177
Qy 181 WSEVQTFRLRLVGLKFLDIPQIOVGLVQYGESPVHWSLGDFTKBEVRAAKNLSRREG 240
Db 178 WVEVQHFLINILKFFYIGPGQIQVGVVQYGEDVHVEFLNDYRSVKDVEAASHIEQRGG 237
Qy 241 RETKTAQIIVACTEGFSQSHGGRPEARLLVVVTDGSHDGEELPAALKACEAGRTRY 300
Db 238 TETRTAFGIFARSEAFQK--GGRKGAKKVMIVITDGSHDSDPLEKVIQOOSERDNVTRY 295
Qy 301 GIAVLGHLRRQDRPSSFLREIRTIASDPDERFFNVTDGSHDSDPLEKVIQOOSERDNVTRY 295
Db 296 AVAVLGYNRRGINPETFLNEIKYIASDDPKHFFNVTDGSHDSDPLEKVIQOOSERDNVTRY 295
Qy 361 HAENESSFGLMSQIGSFTHRLKDGILFGMVAGYDWCSSVLMWEGGHLRFPFRMALEDEF 420
Db 356 N-KNETSFGLMSQIGSFTHRLKDGILFGMVAGYDWCSSVLMWEGGHLRFPFRMALEDEF 420
Qy 421 PPALONHAAYLVGSVSMLLRGGRRLFLSGAPRFRHKGKVIATOLKKGAVRVAOSLOGE 480
Db 415 PEEKLNHGAYLVGTVTSVSSRQGRVYVAGAPRFRHKGKVIATOLKKGAVRVAOSLOGE 480
Qy 481 QIGSYFGSEITSDVIDGQVTDVLLVAGAPMYFN--EGRERKVVYVEL--RQNRFYNGTLK 532
Db 475 QIGSYFGSEITSDVIDGQVTDVLLVAGAPMYFN--EGRERKVVYVEL--RQNRFYNGTLK 532
Qy 541 PEPP-QDARFGFAMGALPDNLQDGFADVAVGAPLEDGHOGALYLYHGQTSGVRPHPAQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNQDSYNDVAVGAPLEDNHAGAIYIFHGRFSILKTPKQRI 592
Qy 600 AASMPHALSYFGSRVDRGLDGLDGLVAVAGAGAAIILSSRPVHLTPSLEVTQAI 659
Db 593 TASELATGLQYFGCSIHGQDLNEDGLIDLAVGALNAVILWSRPVQINASHLFEPSKI 652
Qy 660 SVVQDCCRREGOEAVCLTAALCFQVTSRTPCRWHDHOFYMRFTASLDGWTAGARAFDGG 719
Db 653 NIFHRDCKSRGRDATCLAAFLCTPIFLAPHFQTTVGIIRYNATMDERRYPTRAHLDEGG 712
```

RESULT 6

US-09-836-353A-35

```
; Sequence 35, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PP489P1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-836-353A-35
```

```
Query Match 39.5%; Score 2344.5; DB 10; Length 1189;
Best Local Similarity 42.2%; Pred. No. 2.3e-198;
Matches 505; Conservative 204; Mismatches 407; Indels 81; Gaps 16;

Qy 1 MELPFTVTHLPLVFLVLTGLCSPPNLDDEHPRLPGPPEAFYGSVLQHVGGQRMVLVGA 60
Db 1 MDLPRLGVAAWALSMPGFTDTFNMDTRKPRVPGSRTAFFGYTVQHQHDSGNKWLVVGA 60
Qy 61 PWGPGSDRGDVRCPVGGAHNAPCAKGLHGLDYQLGNSHPAVNMHLGSLLETGDDGG 120
Db 61 PLETNGYQKTDVYKCPV---IHGNCCKLNLGRVTLNVSERKDNMRLGLSLATNPKDNS 117
Qy 121 FMACAPLWSRACSSVSSFGICARVDASFOQSGSLAPTAQRCPTYMDVIVLDGNSIYP 180
Db 118 FLACPLWSEHCSSYYTTCMCSRVSNNFRFSKTVAPALQRCQTYMDIVIVLDGNSIYP 177
Qy 181 WSEVQTFRLRLVGLKFLDIPQIOVGLVQYGESPVHWSLGDFTKBEVRAAKNLSRREG 240
Db 178 WVEVQHFLINILKFFYIGPGQIQVGVVQYGEDVHVEFLNDYRSVKDVEAASHIEQRGG 237
Qy 241 RETKTAQIIVACTEGFSQSHGGRPEARLLVVVTDGSHDGEELPAALKACEAGRTRY 300
Db 238 TETRTAFGIFARSEAFQK--GGRKGAKKVMIVITDGSHDSDPLEKVIQOOSERDNVTRY 295
Qy 301 GIAVLGHLRRQDRPSSFLREIRTIASDPDERFFNVTDGSHDSDPLEKVIQOOSERDNVTRY 295
Db 296 AVAVLGYNRRGINPETFLNEIKYIASDDPKHFFNVTDGSHDSDPLEKVIQOOSERDNVTRY 295
Qy 361 HAENESSFGLMSQIGSFTHRLKDGILFGMVAGYDWCSSVLMWEGGHLRFPFRMALEDEF 420
Db 356 N-KNETSFGLMSQIGSFTHRLKDGILFGMVAGYDWCSSVLMWEGGHLRFPFRMALEDEF 420
Qy 421 PPALONHAAYLVGSVSMLLRGGRRLFLSGAPRFRHKGKVIATOLKKGAVRVAOSLOGE 480
Db 415 PEEKLNHGAYLVGTVTSVSSRQGRVYVAGAPRFRHKGKVIATOLKKGAVRVAOSLOGE 480
Qy 481 QIGSYFGSEITSDVIDGQVTDVLLVAGAPMYFN--EGRERKVVYVEL--RQNRFYNGTLK 532
Db 475 QIGSYFGSEITSDVIDGQVTDVLLVAGAPMYFN--EGRERKVVYVEL--RQNRFYNGTLK 532
Qy 541 PEPP-QDARFGFAMGALPDNLQDGFADVAVGAPLEDGHOGALYLYHGQTSGVRPHPAQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNQDSYNDVAVGAPLEDNHAGAIYIFHGRFSILKTPKQRI 592
Qy 600 AASMPHALSYFGSRVDRGLDGLDGLVAVAGAGAAIILSSRPVHLTPSLEVTQAI 659
Db 593 TASELATGLQYFGCSIHGQDLNEDGLIDLAVGALNAVILWSRPVQINASHLFEPSKI 652
Qy 660 SVVQDCCRREGOEAVCLTAALCFQVTSRTPCRWHDHOFYMRFTASLDGWTAGARAFDGG 719
Db 653 NIFHRDCKSRGRDATCLAAFLCTPIFLAPHFQTTVGIIRYNATMDERRYPTRAHLDEGG 712
```

```
Qy 720 QRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALTFTFALDNTTKPGPVLNEGSPTSI 779
Db 713 DRFTNRAVLSSGQELCERINFHVLDTADYVKVPTFVSYSLEDPDH-GPMLDDGWPTTL 771
Qy 780 QKLVPPSKDCGPNCEVTDLVLQVMDI-----RGRK-----APFVV 817
Db 772 RVSPVFWGNCNEDSHCVPLDLVARSDLPTAMEYQORVLRKPAQDCSAYTSLSDFTTVFII 831
Qy 818 RGRKRVLTSTLENKENAYNTSLIFSRNLHLASLTQORSPPIKVECAAPS--AHAR 875
Db 832 ESTRQRAVATLENRENAYSTVINISQSANQFASLIQKEDSDGSEICNEERLQKQ 891
Qy 876 LCSVGHVPVOTGAKVTFLLEFPFSCSLLSQVFGKLTASSDSLSRNGTLOENTTAQTSAYI 935
Db 892 VCNVSYPFRAKAKVAFRLDFEFSKSIPLHLELELAAGSDSNERDSTKEDNVAFLRFL 951
Qy 936 QYBPHLLFSSESTLHRVEVHPYGTLP--VGPGBPFKTLRTN----- 975
Db 952 KYEADVLFTRSSLSHYEVLNSSLERYDGGIGPPFCIFRIQNLGLFPIHGMKTIPI 1011
Qy 976 -----NASC-IVQNLTETPPGPPVHPPEELOHTNRLNGSNTQCQVR 1014
Db 1012 ATRSGNRLKLRDLTDEVANTSCNIWNGSTERYPTVE-EDURRAPQLAHNSDVVVSIN 1070
Qy 1015 CHLQQLAKGTEVSVGLLRLVNEPFRRAKFKSLTVVSTFELGTBEGSVLQ/LTEASRWSES 1074
Db 1071 CNI-RLVPNQEPHLLGNLWLRSLKALKYKSMKIMYNAAALQRFHSPFIFREEDPSRQI 1129
Qy 1075 LLEVQTRPILISWILIGSVLGGILLALLVFCWLKLGFFAHKAIPEEKREBKELE 1131
Db 1130 VFEISKQEDMQVPTIIVGSLTGLGULLLALLVLAALWKLGFPR SAR----RRREPGLD 1182

RESULT 7
US-10-262-839-4
/ Sequence 4, Application US/10262839
/ Publication No. US20040038877A1
/ GENERAL INFORMATION:
/ APPLICANT: Alsebrook, David,
/ APPLICANT: Anderson, John W.,
/ APPLICANT: Boldog, Ferenc,
/ APPLICANT: Burgess, Catherine,
/ APPLICANT: Catterton, Elina,
/ APPLICANT: Edinger, Shlomit,
/ APPLICANT: Ellerman, Karen,
/ APPLICANT: Gerlach, Valerie,
/ APPLICANT: Gorman, Linda,
/ APPLICANT: Ji, Weizhen,
/ APPLICANT: Kekuda, Rameah,
/ APPLICANT: Leach, Martin,
/ APPLICANT: Li, Li,
/ APPLICANT: Miller, Charles,
/ APPLICANT: Patturajan, Meera,
/ APPLICANT: Reiger, Daniel,
/ APPLICANT: Rothenberg, Mark,
/ APPLICANT: Shinkets, Richard,
/ APPLICANT: Smithson, Glenda,
/ APPLICANT: Spytek, Kimberly,
/ APPLICANT: Tautier, Raymond, Jr.,
/ APPLICANT: Vernet, Corine,
/ APPLICANT: Voss, Edward,
/ APPLICANT: Zerhusen, Brian,
/ APPLICANT: Zhong, Mei
/ TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
/ FILE REFERENCE: 21402-462A
/ CURRENT APPLICATION NUMBER: US/10/262,839
/ CURRENT FILING DATE: 2002-10-01
/ PRIOR FILING DATE: 2002-10-01
/ PRIOR FILING DATE: 2001-10-02
/ PRIOR FILING DATE: 2001-10-09
```

```
/ PRIOR APPLICATION NUMBER: 60/328,029
/ PRIOR FILING DATE: 2001-10-09
/ PRIOR APPLICATION NUMBER: 60/328,056
/ PRIOR FILING DATE: 2001-10-09
/ PRIOR APPLICATION NUMBER: 60/381,101
/ PRIOR FILING DATE: 2002-05-16
/ PRIOR APPLICATION NUMBER: 60/371,972
/ PRIOR FILING DATE: 2002-04-12
/ PRIOR APPLICATION NUMBER: 60/327,342
/ PRIOR FILING DATE: 2001-10-05
/ PRIOR APPLICATION NUMBER: 60/328,044
/ PRIOR FILING DATE: 2001-10-09
/ PRIOR APPLICATION NUMBER: 60/328,849
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: 60/374,738
/ PRIOR FILING DATE: 2002-04-23
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 367
/ SOFTWARE: Curaseq1ist version 0.1
/ SEQ ID NO 4
/ LENGTH: 1189
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-10-262-839-4

Query Match 39.5%; Score 2344.5; DB 15; Length 1189;
Best Local Similarity 42.3%; Pred. No. 2.3e-198;
Matches 505; Conservative 204; Mismatches 407; Indels 81; Gaps 16;

Qy 1 MELPFTVTHLPLVFLVTGLCSPPNLDHHPRLPPGPEABFGYSVLVHGVGGORWMLVGA 60
Db 1 MDLPRLGVVAWALSMLPFGFTDTFNMMDTRKPRVIPGSRTPAFGYTVQOHDISGNKWL VVGA 60

Qy 61 PWDGPGDREGDYVRCVPVCGAHNAPCAKGLHGDYQIQLGNSSHPAVNMHLGMSLLETDCGG 120
Db 61 PLETNQYQKTGDYKCPV-----IHGCTKMLGRVTLNSVSRKDNMRGLGSLATNPKNDS 117

Qy 121 FMACAPLWSRACSSVFSFGICARVDASFOQGS LAPTAQRCPTYMDVTVLVDGSNSIYP 180
Db 118 FLACSPMSHECGSSVYTCMCSRNSNFRFSKTVAPALQRCQTYMDIVLVDGSNSIYP 177

Qy 181 WSEVQTFRLRLVKLFIDPEQIOVLGVQYESPVHWSLGDFTKEBVEVVAANKLSRREG 240
Db 178 WVEVQHFLINILKKFYIGPQIQVGVVQYGEDVHFEHLNDYRSVDVVEAASHIEQRG 237

Qy 241 RETKTAQIMVACTEGFSQSHGGRPEARLLVVVTGESHGDEELPAALKACEAGRVTRY 300
Db 238 TETRTAFGIEFARSEAFQK--GGRKGAKYMI VITGESHSDSPLEKVIQQSERKDNVTRY 295

Qy 301 GIAVLGHYLRQRDPSPFLREIRTIASDPDERPFFNVNTDEAALTDI VDALGDRIFGLGEGS 360
Db 296 AVAVLGYNNRRGINPETFLNEIKYIASDDPKHFFNVNTDEAALKDI VDALGDRIFSLEGT 355

Qy 361 HARNESSFGLEMSQIGFSTHRLKDGILFGMVGYDMOGSVLMLEGGHRLFPFRMALEDEF 420
Db 356 N-KNETSFGLEMSQTGFSHVHVEDGVLLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEF 414

Qy 421 PPALQNHAAVLYGYSVSSMLLGGRRLLSLGAPRFRHKGKVIAPOLKKGAVRVAQSLQGE 480
Db 415 PEELKNHGAYLGYTVTVSVSSRQGRVYVAGAPRPNHTGKVLFTFMHNNRSLTIHQAMRGG 474

Qy 481 QIGSYFGSELCPDLTDRDGTDTDLVLLVAA PMFLGAPQNKETGRVYVYLVGQOQSLTLQGTTLQ 540
Db 475 QIGSYFGSEITSDVIDDGDGVTDLVVGAPYFN-EGRERKGVYVEL-RQNRVYVNGTLK 532

Qy 541 PEPP-QDARTFGAMGALPDNLNQGFADVAGAPLEDHQGALYLYHGTQSGVRPHPAQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNQSDYNDVVVVGAPLEDNHAGAIYIFHGFSGSILKTPKQRI 592

Qy 600 AAASMPHALSVRGSVDGLDLDGDDLDVAVGAGQAAILLSRPIVHLTPSLEVTPOAI 659
Db 593 TASELATGLQYFGCSIHGQDLNEDGLIDLAVGALGNVILMSRPVVQINASLHFEPSKI 652
```

[illegible]

Query Match 37.0%; Score 2194; DB 10; Length 1034;  
Best Local Similarity 44.3%; Ered. No. 4.3e-185;  
Matches 460; Conservative 178; Mismatches 346; Indels 54; Gaps 13;

Qy	61	PWDGPGDRRGDYVRCVGGAHNAPCAKGHLDGYQI GNSHPAVNMHLGNSLILETDGCG	120
Db	61	PLETNGYQKTGDYVKCPV---IHGNCITKLNLRVTLNSVSEKDNKRNRLGLSLATNPKNDS	117
Qy	121	FMACAPLWSRACSSYVSSGICARVDASFOPOGSLAPTAQRCPTYMDVIVLDCGNSIYP	180
Db	118	FLACSPWHSCECSSYYTGMCSRVSNSNFFSKTVAPALQRCQTYMDIVIVLDCGNSIYP	177
Qy	181	WSBVQTLRLVOKLFDIPBQIOGLVVOYGESPVHEWSLGDFTKTEVVRAAKNLSRREG	240
Db	178	WVBVQHLINILKKFVGPQIOGVVVOYGEDVVHFLNDYRSVKDVEAAASHIEQORG	237
Qy	241	REYTKTAQIMVACTEFSQSHGRPEARALLVVVTDGESHDBEELPAALKACACAGRVTRY	300
Db	238	TETRTAPGIEFAFSEAFQK--GGRKGAKKVVITDGESHDPDEKVIQOOSERNVTRY	295
Qy	301	GI AVLGHYLRQRDPSPSFLREIRTIASDPDERFEFFNVTDAAALTDIVDALGDRIFGLGEGS	360
Db	296	AVAVLYGYNPERGINPETFLNEIKYIASDPDDKHFFNVTDAAALKDVIDALGDRIFSLGEGT	355
Qy	361	HAENESSFGLEMSQIOGFSTHRLKDGILFGMWGDAYDWGGSVLWLEGGHRLPFPRLMALBDEF	420
Db	356	N-KNETSFGLEMSQTFSSHVVEDGVLLGAVGAYDMNGAVLKETSAKVPLRESYLKEF	414
Qy	421	PPALQNHAAVLYGYSVSSMLLRGRRLLFGSAPRFRHGKVIAFOLKXDGAVRVAQSQGE	480
Db	415	PEELKHGAVLGTIVTISVSSROGRVTVAGAPRNTGKVLFTMHNRSLSUTHQANRGQ	474
Qy	481	QIGSYFGESECLPDTDRDGTVDLLVAAPMFLGPQNKETGRVYVYLVGQOSSLTLQSTLQ	540
Db	475	QIGSYFGESEITSDIDGCVTDVLLVGAPVFN--EGREGRKVYVEL--RQNRVFNVTGLK	532
Qy	541	PEPP--QDARFGPANGALPDNLQDGFADVAVGAPLEDHQCALVLYHGTQSGVRPHPAQRI	599
Db	533	DSHSYQNAREFGSIAVRDLNQDSYNDVVVVGAPLEDNHAGAIYIFHGFSGSLKTPKQRI	592
Qy	600	AAASMPHALSYFGRSVDRGLDLDGDDLVDVAVAGQAAGAAI LSSRPVHLTPSLEVTPOAI	659
Db	593	TASELATGLQYFGCSIHGQLDLNEDGLIDLAVGALGNVILWSPPVQIINASLHFEPSKI	652
Qy	660	SVVQDRCRRGQBAVCLTAALCFQVTSRTPGRWDHQFYMRFTASLDEWTTAGARAAFQSG	719
Db	653	NI FHRDKCRGRDATCLAAFLCPTPIAPHFQTTVGI RYNATMDKRYTPPRAHLDEGG	712
Qy	720	QRLSPRLRLSVGNVCEQLHFHVLDTSDYL RPVALVTPEALDNTTKPGPVNLGSGTSL	779
Db	713	DRFTNRAVLLSSGQELCERINFHVLDTADYVKPVTFSVYSLEDPDH--GPMLDGDPWPTTL	771
Qy	780	QKLVPPSKDCGPONECVTDLVLQVNMDI-----RGSRK-----APFVV	817
Db	772	RVSVPFWNGNEBDEHCVPDLVLDARSDLPTAMEVCQVRVKRPAQDCSAYTSLSPDTVPFII	831
Qy	818	RGRRKVLVSTTLNKRKENAYNTSLSTIFSRNLHLASLTPORESPIKVECAAPS--AHAR	875
Db	832	ESTRQRVAVEATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSEICVNEERRLQKQ	891
Qy	876	LCSVGHEVPQTAGKVTLLBFEFSCSSLLSQVEFKLTASSDSLERNCTLOENTAQTSAYI	935
Db	892	VCMVSYPPFFRAKAVAFRLDPFEFSKIFLHHLETELAAGSDSNERSDKEDNVAPLPHL	951
Qy	936	QYBPHLLFSSSESTLHYRVEHPYGTLP--VGPGPEFKTLTRTNNASCTVQNLTPEPGPVH	993
Db	952	KYEADVLFTRSSSLSHYEVKLNLSRLERYDGI GPPFFSCIFRIQNLGLF-----PIH	1001
Qy	994	-----PEELQHTNRL 1003	
Db	1002	GIMMKITIPIATRSGNRL 1019	

RESULT 9  
US-09-836-353A-43  
; Sequence 43, Application US/09836353A  
; Publication No. US20030129685A1

US-09-984-130-43	US-09-984-130-43	RESULT 8	
; Sequence 43, Application US/09984130			
; Publication No. US20030055231A1			
; GENERAL INFORMATION:			
; APPLICANT: Ni et al.			
; TITLE OF INVENTION: 12 Human Secreted Proteins			
; FILE REFERENCE: PF489P2			
; CURRENT APPLICATION NUMBER: US/09/984,130			
; CURRENT FILING DATE: 2001-10-29			
; PRIORITY APPLICATION NUMBER: 60/243,792			
; PRIORITY FILING DATE: 2000-10-30			
; PRIORITY APPLICATION NUMBER: 09/836,353			
; PRIORITY FILING DATE: 2001-04-18			
; PRIORITY APPLICATION NUMBER: 60/198,407			
; PRIORITY FILING DATE: 2000-04-19			
; PRIORITY APPLICATION NUMBER: PCT/US99/25031			
; PRIORITY FILING DATE: 1999-10-27			
; PRIORITY APPLICATION NUMBER: 60/105,971			
; PRIORITY FILING DATE: 1998-10-28			
; NUMBER OF SEQ ID NOS: 149			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 43			
; LENGTH: 1034			
; TYPE: PRMT			
; ORGANISM: Homo sapiens			



```

; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: P489P1
; CURRENT APPLICATION NUMBER: US/09/836.353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR FILING DATE: 2001-04-18
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-353A-43

Query Match 37.0%; Score 2194; DB 10; Length 1034;
Best Local Similarity 44.3%; Pred. No. 4.3e-185;
Matches 460; Conservative 178; Mismatches 346; Indels 54; Gaps 13;

Qy 1 MELPFVTHLFLPLVFLTGLCSPPFLDEHHRLFPGLPPEABFGYSLVQLHVGQGRWMLVGA 60
Db 1 MDLPRLGVAAWALSLWPGFTDTFNMTRKPRVIFGSRTPAFGYTVQQHDISGNKWLWVGA 60
Qy 61 PWDPGSDRGDVRCPVGGNAHPCAKGHLDGLQNGSHHPAVNMHLGMSLLETDDGG 120
Db 61 PLETFNGVQKTDVYKCPV--IHGNCCKNLNIGRVTLSNVSRKDNMRLGLSLATNPKNDS 117
Qy 121 FMACAPLWSRACGSSVSSGICARVDASFOQSGSLAPTAQRCPTVMVVLVDSGNIYP 180
Db 118 FLACSPINSHCEGSSYTTGCSRNVSNFRPSKTVAPALQRCQTYMDIVVLDSGNIYP 177
Qy 181 WSEVQTLRLRLVGLKFLIDPEQIQVLQVYGSPPVHWSLGDPRTKBEVVRANKLSRREG 240
Db 178 WVEVQHLINILKKFIYGPQIQGVYQVQGVDDVHFLNDYRSVKDVEAASHIEQGG 237
Qy 241 RETTAQAINVACTEGSQSHSGRPEARLLVVTVDSHSGDELPALAKACEAGRTRY 300
Db 238 TETATFAGIEFARSEAFQK--GGRKGAKVMIVITDGHSDSPDLEKVIQOOSRDNVTRY 295
Qy 301 GIAVLGHLRQRDPSSFLRIRTIASDPDERFFNVNTDEAALTDIVDALGDRIFGLEGS 360
Db 296 AVAVLGYNRGINPEFLNEIKVIASDPDKHFENVTDDEAALXDIVDALGDRIFSLGEGT 355
Qy 361 HAENESSFGLEMSQIGFSTRHLKDKGILFGMVGAYDWGQSVLWLEGHRLPFPFRMALEDEF 420
Db 356 N-KNETSGFLEMSQTGFSSHVVEDGVLLGAVGYDMNGAVLKETSAGKVIPLRESYLKEF 414
Qy 421 PPALONHAAYLYGSVSSMLLGGRLFLSGAPRRFRHKGKVIAPOLKDGAVRVAQSLQGE 480
Db 415 PEELKNHAGLYGTVTSVSSRQGRVYVAGAPRNHTGKVLFTMHNNSLTIHQAMRGQ 474
Qy 481 QIGSYFSGELCPDLDRDGTDLVLLVAMPFLGPQNKETGRVYVYLVQVQOQSLTLQGTLO 540
Db 475 QIGSYFSGEITSDVLDGGVTDVLLVAGPMYFN--EGRERGKYYVEL-RQNRFYVNGTLK 532
Qy 541 PEPP-QDARFGFAMGALPDNLQDGFADVAVGAPLEDHQGLYLYHGTSQGRVPHPAQRI 599
Db 533 DSHSYQNARFGSSIASVRDLNODSYNDVVVVGAPLEDNHAGAIYIFHGRGSIKTKPQRI 592
Qy 600 AAASMPHALSVFGRSDRLDGDGLVDVAVGAQCAAILLSRPIVHLTPSLVTPQAI 659
Db 593 TASELATGLQYFGCSIHQQLDNEGLDIDLAVGALGNVILWSRPVQVQINASLHFEPSKI 652
Qy 660 SVQRDCRRRGOEAVCLTAALCFQVTSRTPGRWDHQFYMRTASLDWTAGARAADFGSG 719
Db 653 NIFRDCRCKSRDATCLAAFLCFTPIFLAFHQTITVIGIRNATMDEKXYTPRAHLEGG 712
Qy 720 QRLSPRLRLSVGNVTCQLHFLHVLDTSDYLRLPVALTVTFALDNTTKPGPVLNEGSPTSI 779
```

```

Db 713 DRFTNRAVLLSSQOELCERINFHVLDTADYVKVETFSVEYSLEDPDH-GPMLDDGWPTTL 771
Qy 780 QKLVPPSKDCGPDNECVTDLVLQVNMDI-----RGSRK-----APFVV 817
Db 772 RVSVPFWNGCNEDEHCVDPDLVLDARSDLPMTAMEYCORVLKPKAQDCSAYTLSPDTTVFII 831
Qy 818 RGGRRKVLVSTTLERKENAYNTSLSIIFSRNLHLASLTPQRESPIKVECAAPS--AHAR 875
Db 832 ESTRQRVAVENTLENRGENAYSTVLNISQSANLQFASLIQKEDSGSIECVNEERRLQKQ 891
Qy 876 LCSVGHVPVOTGAKVTFLLFEPSCSLLSQVFGKLTASSDSLSRNGTLOENTAQTSAIYI 935
Db 892 VCNVSYPPFFRAKAKVAFRLDFFESKSIFLHLEILAAGSDSNERSDSTKEDNVAPLRFHL 951
Qy 936 QYEPHLLFSSESTLHRYEVHPYCTLP--VCPGPEEKTTLRTNNASCIVQLNLTSPGPPVH 993
Db 952 KYEADVLFTKSSLSLHYEVKLNSSLERYDGIQGFPPCFRIQNLGLP-----PIH 1001
Qy 994 -----PEELQHTNRL 1003
Db 1002 GIMMKITIPATRSNRL 1019

RESULT 10
US-10-262-839-6
; Sequence 6, Application US/10262839
; Publication No. US20040038877A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John,
; APPLICANT: Anderson, David W.,
; APPLICANT: Boldog, Ferenc,
; APPLICANT: Burgess, Catherine,
; APPLICANT: Catterton, Elina,
; APPLICANT: Edinger, Shlomit,
; APPLICANT: Ellerman, Karen,
; APPLICANT: Gerlach, Valerie,
; APPLICANT: Gorman, Linda,
; APPLICANT: Guo, Xiaojia,
; APPLICANT: Ji, Weizhen,
; APPLICANT: Kekuda, Ramesh,
; APPLICANT: Leach, Martin,
; APPLICANT: Li, Li,
; APPLICANT: Miller, Charles,
; APPLICANT: Patturajan, Meera,
; APPLICANT: Reiger, Daniel,
; APPLICANT: Rothenberg, Mark,
; APPLICANT: Shimkets, Richard,
; APPLICANT: Smithson, Glenda,
; APPLICANT: Spytek, Kimberly,
; APPLICANT: Taupier, Raymond, jr.,
; APPLICANT: Vernet, Corine,
; APPLICANT: Voess, Edward,
; APPLICANT: Zernusen, Brian,
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS
; FILE REFERENCE: 21402-462A
; CURRENT APPLICATION NUMBER: US/10/262,839
; CURRENT FILING DATE: 2002-10-01
; PRIOR FILING DATE: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,101
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/371,972
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/327,342
; PRIOR FILING DATE: 2001-10-05
```





```
Qy 255 EGFSSQSHGGRPEARLLVVTGESHGDBELPAALKAACEAGRVTRYGIAVLGHYLRORD 314
Db 260 EAFTEARGARRGVKKVMVITDGHSHNHLKKVQDCEENIQRFSAIILGYSYNRGNLS 319
Qy 315 PSFRLIRITIASDPRFPNVTDEAALTDIVDALGDRIFGLEGSHAENESSFGLMSQ 374
Db 320 TEKFVEBIKSIASPTKEKFFNVSDDELALVTIVKTLGERIFALEATADQSAASFEMMSQ 379
Qy 375 IGFSRHLKDGILFGMVGAYDMGSGVLWEGGHRLLPPRMALEDEFPFALQNHAAVLGYS 434
Db 380 TGFSAHYSQDWMLGAVGADWNGTVVMQASQIIIPNTTFNVESTKNEPLASYLGYT 439
Qy 435 VSSMLLGGRRLLFUSGAPRFRHRGKVIAFOLKDGAVRAQSLOGEQIGSYFSGELCPD 494
Db 440 VNSATSSGDVLYTAGQPRYNHTGOVVIYRM-EDGNIKILQTLGSEIGSYFSGILTTD 498
Qy 495 TDRGTTDVLVVAAPMFLGPNKQKTRGVYVYLVCQQLLTQGLQP----- 541
Db 499 IDKDSNTDILLVGAPMYMGTEKEGKGVYVYAL-NOTRFEYQMSLEPIKOTCCSSRQHNS 557
Qy 542 -----BPPQDARFGFANGALPDNLQDGFADVAGAPLEDHGHQALYLYHGTQSGVRPH 595
Db 558 CTTEKNEPFCARPTAIAAVKDLNLDGFNDIVIGAPLEDHGHGAVYIYHSGSKTIRKEY 617
Qy 596 AQRTAAASMPHALSYFQSRVDGLDGLDGLDVAVAGAQGAAILSSRPVHLTPSLEVT 655
Db 618 AQRIPSGGDGKTLKFFGOSIHGENDLNGDGLTDVTIGLGAALFMSRDVAVVKVWTFE 677
Qy 656 PQAISVVQDRCRRQCAVCLTAALCFQVTSRTPGRWDPHQPMPFTASLDEWTAGARAA 715
Db 678 PNKVNIOKKCHMEGKETVCINATVCPVEVLSKSKEDTIYEADLQRYVTLDSLRSRFF 737
Qy 716 DGSQRLSPRLRLSVGNVTCEQLHPHVLDTSDYLRPVALVTALDNTTKP--GPVLNE 773
Db 738 SGTQERKQVR--NITVRKSECTKHSFYMLDKHDFQDSVRITLDP---NLTDPENGPVLD 792
Qy 774 GSPTSIOKLVPFSCDGPDNBECVTDVLQVNMIDIRSKAPFVVRGGRKRVLVTTLNLR 833
Db 793 SLPNVHEIYIPFAXDCGKKEKICISLSLHV---ATTEKDLIIVRSQNDKFNVSILT 848
Qy 834 KENAYNTSLIIFSRNLHLASLTQRESPIKVECAAPSAHARLCSVGHVPFQCAKUTFL 893
Db 849 KDSAYNTRTIYHSPNLVFSGI-----EAIQKDSK--ESNHNITCKVGYPFLRRGEM 902
Qy 894 LEPFSCSSLLSQVFGKLTASSDSLERNGLTQENTAOISAVIOYEPHLLPSESSETHRYE 953
Db 903 ILFQNTSYLMENVTIYLSATSDSEEPETLSDNVVAISIPVKYEVGLQFYSSASEYHIS 962
Qy 954 VHPYGTLP-----VG-----PQPEFKTTLRTNNASCIVQNLTPEPPGP 990
Db 963 IAAENTVPEVINSTEDIGNEINIFYLIRKSGSPMPPELKLIS-----PPNMTSNGYP 1015
Qy 991 PVHPEELQHTNRLN-----GSTQCO-----VVRCHLG 1018
Db 1016 VLYPTGLSSSENACRPHIPEDPFSINSKGKMTTSTDHLKRGTTLDNCNCKFATITCNLT 1075
Qy 1019 QLANGTEVSGLLRLVNERPRRAKFKSLTVVSFELGTBEGSVLQLTEASRWSSESLLEV 1078
Db 1076 S-SDISQVNSL--ILWKPTFKSYFSSNLNITIGELRSENAS-LVLSSSQKKELAIQI 1131
Qy 1079 VQT-RPLISILWILGVLGLLALLAVFLWLKGFPAHKKIPEEKREE 1128
Db 1132 SKDGLPGRVPLVILLSAFAGLLMLLILALWKIGFF---KRPKKKMEK 1179
```

## RESULT 12

```
US-09-918-715-307
; Sequence 307, Application US/0918715
; Publication No. US20030017157A1
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Bert Vogelstein
; APPLICANT: Kenneth Kinzier
```

```
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
; CURRENT APPLICATION NUMBER: US/09/918,715
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/222,599
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 60/224,360
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 307
; LENGTH: 1180
; TYPE: PRT
; ORGANISM: Rat
; US-09-918-715-307

Query Match 31.8%; Score 1884; DB 10; Length 1180;
Best Local Similarity 36.2%; Pred. No. 2e-157;
Matches 432; Conservative 219; Mismatches 432; Indels 112; Gaps 25;

Qy 13 LVFLTGLCSPPNLDHHPLRFPQPPPEASFGYSVLQHVGGQRMVLGAPWDGSGDRGD 72
Db 19 LTVILGFCVSNVDVKNMSFSFGFVEDMFGYTVQYENEEGKRWLIGSLVLPVGPQKARTGD 78
Qy 73 VYRCPVGAHNAFCAGKHLGDYQLGHSNHPAV----NMHLGMSLLETDCDGGCFMACPL 127
Db 79 VYKCPVGRERAMFCVKLDLP----VNISIPNVTEIKENMTFG-STLVNPNGGFLACGPL 133
Qy 128 WSRAGSSVFSFGICARVDASFPQGS LAPTAQRCPTYMDVIVLDGNSIYPMSEVQTF 187
Db 134 YAYRCGHLHTTGICSDVSPTFQVNSFAP-VQECSTQLDIVLDGNSIYPMSEVIAF 192
Qy 188 LRLVGLKFLDPQIQVGLVQYGESPHWESLGDFTKEBVRVRAAKLSRREGRETTAQ 247
Db 193 LNDLLKRMIDIGPQTQGVIGYGENVTHEFNLNKYSSTEELVAANKIGRGGQLNTMAL 252
Qy 248 AIMVACTGFGSQSHGGRPEARLLVVTGESHGDBELPAALKAACEAGRVTRYGIAVLGH 307
Db 253 GIDTARKEATEARGARGVKVMVITDGHSHNHLKQVQDCEENIQRFSAIILGH 312
Qy 308 YLRQRDPSSFLREIRTIASDPRFPNVTDEAALTDIVDALGDRIFGLEGSHAENESS 367
Db 313 YNRGNLSTEFVSEIKSIASEPTKEKFFNVSDDELALVTIVKTLGERIFALEATADQSAAS 372
Qy 368 FGLMSQIGFSTRHLKDGILFGMVGAYDMGSGVLWEGGHRLLPPRMALEDEFPFALQNH 427
Db 373 FEMMSQTGFSAHYSQDWMLGAVGAYDMNGTVVMQANQMVIPHNTTFQTE--PAKONE 430
Qy 428 --RAYLGYSYSSMLLRRGRRLFLSGAPRFRHRGKVIAFOLKDGAVRAQSLOGEQIGSY 485
Db 431 PLASYLGYTVNSATIPGD-VLYTAGQPRYNHTGOVVIYKM-EDGNINILQTLGGEQIGSY 488
Qy 486 FGSGLCPDTRDGTDDVLVAAAPMFLGPNKQKTRGVYVYLVCQQLLTQGLQP---- 541
Db 489 FGSVLTITIDKDSYTDLLLVGAPMYMGTEKEGKGVYVYAL-NOTRFEYQMSLEPIRQT 547
Qy 542 -----EPPQDARFGFANGALPDNLQDGFADVAGAPLEDHGHQALYLYHG 586
Db 548 CCSKLDNCSCTKENKNEPFCARPTAIAAVKDLNLDGFNDIVIGAPLEDHGHGAVYIYHG 607
Qy 587 TQSGVRPHPAQRTAAASMPHALSYFGRSDVGRDLDDGLDGLDVAVAGAQGAAILSSRPV 646
Db 608 SGKTIREAYAQRIPSGGDGKTLKFFGOSIHGENDLNGDGLTDVTIGLGAALFWARDVA 667
Qy 647 HLTPSLEVTQPAISVVQDRCRRRQCAVCLTAALCFQVTSRTPGRWDPHQPMPFTASLDE 706
Db 668 VVKVTMNFENKVNIOKKCHMEGKETVCINATVCPVEVLSKSKEDSIYEADLQRYVTLDS 727
Qy 707 WTAGARAAFDGSGQRLSPRLRLSVGNVTCEQLHPHVLDTSDYLRPVALVTTFALDNTTK 766
Db 728 LRQISRSFFSGTQERKQVR--NITVRSEBCIRHSFYMLDKHDFQDSVRVITLDP---NLTD 782
```



```
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR FILING DATE: 2001-04-18
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 103
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-836-353A-103

Query Match      31.7%; Score 1881.5; DB 10; Length 1151;
Best Local Similarity 35.9%; Pred. No. 3.2e-157;
Matches 426; Conservative 220; Mismatches 423; Indels 119; Gaps 23;

Qy 23 FNLDEHPRLFPFGPPEAEFGYSLVQHVGGQRMWLVGAPMDPGSDRRGDYVRCVGGAH 82
Db 1 FNVVDKNSMTFSGPVEDMFGYVQYENEBGKWLIGSLVGPQKNRTGDYVCKPVRGE 60
Qy 83 NAPCAKGLHDYQLGNSSHPAV-----NMHLGMSLLETDGDDGGMACAPLWSRACGSSVF 137
Db 61 SLPCVKLDLP-----VNTSIPNVTBVENMTFG-STLVTPNPNNGFLACGPLYAYRCGHLHY 115
Qy 138 SSGICARVDASFQPGQSLAPQRCPTVMDVIVLDGNSNIYPWSEVQTFRLRLVGLFI 197
Db 116 ITGICSDVSPFTQVNSIAP-VQECSTQLDIVILDGNSNIYPWSDVTFAPLNDLKKMDI 174
Qy 198 DPEIQIQLVQYSGSPVHNSLGDFTKEEVVRAAKNLSRRREGRETCTAQAIMVACTEGF 257
Db 175 GPKQTQVGIQYQGENVTHEFNLNKYSSTEEVLVAAKKIVQRGGRQCTWTALGTDTRKEAF 234
Qy 258 SQSHGGRPEAARLLVVTDGSHSGEELPAALKACEAGRVTRYIAVLGHYLRQRDPSS 317
Db 235 TEARGARRGVKKVMVITDGSNDHRLKKVIOQCEDENIQRFSAIILGYSYNRGNLSTEK 294
Qy 318 FLREIRTIASDPDRFPFNTVDEALTDIVDALGDRIFGLEGSHAENESSFGLMSQIGF 377
Db 295 FVEIKSIASEPTEKHFNFVSDLEALVTIVKTLGERIFALEATADQSAASFEMMSQTGF 354
Qy 378 STHRLKDGILFGMVGAYDWGGSVLWLEGGHRLFPFRMALEDEFPFALQNHAAVILGYSVSS 437
Db 355 SAHYSQDWMLGAVGADWNGTVVMQKASQIIIPNITTFNVESTKQNEPLASYLGYTVNS 414
Qy 438 MLLRGRRLLFUSGAPRFRHRKVIAPQLKXGDGAVRVAAQSLOGEIGSYFSGELCPDTR 497
Db 415 ATASSGDVLYIAGQPRYNHTQVIIYRM-EDGNKILQTLGSEQIGSYFSGSILTTDIDK 473
Qy 498 DGTITDVLVAAPMLFQNKETGRVYVYLVQCOSSLTTLQGLTLP----- 541
Db 474 DSNTDILLVGAAPMYNGTEKEBQGVYVYAL-NQTRFYEQMSLEPIKQTCSSRQHNSCTT 532
Qy 542 ---BPPQDARFGFANGALPDNLQDGFADVAGAPLEDGHQCALYLYHGTQSGVPHPAQR 598
Db 533 ENKNEPCGARGTAIAAKVLDNLGDFNDIVGAPLEDHGHGAVIYHSGKTIKRYEQAQR 592
Qy 599 IAAASMPHALSYFGRSVDRLDLDGDDLDVVDVAGQAAILLSRPIVHLFPSEVTPQA 658
Db 593 IPSGGDGKTLFFGQSINGEDLNGDGLTDTVIGGLGAALFWSRDAVAVVKNFSPNK 652
Qy 659 ISVVDRCRRRGQEAVALCTAALCFQVTSRTFGRWDHQFYMEFTASLDEWTAGARAAFGS 718
Db 653 VNIQKNCHEGKETVCINATVCFEVKLSKEDTIYEADLQYRVTLDSLRLQISRSFFSGT 712

; Publication No. US10741600
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1161
; LENGTH: 1177
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-741-600-1161

Query Match      31.6%; Score 1874.5; DB 17; Length 1177;
Best Local Similarity 35.9%; Pred. No. 1.4e-156;
Matches 427; Conservative 219; Mismatches 424; Indels 121; Gaps 24;

Qy 20 CSPNLDEHPRLFPFGPPEAEFGYSLVQHVGGQRMWLVGAPMDPGSDRRGDYVRCVPG 79
Db 26 CVSFNVDRKNSMTFSGPVEDMFGYVQYENEBGKWLIGSLVGPQKNRTGDYVCKPVG 85
Qy 80 GAHNAPCAKGLHDYQLGNSSHPAV-----NMHLGMSLLETDGDDGGMACAPLWSRACGS 134
Db 86 RGEISLPCVKLDLP-----VNTSIPNVTBVENMTFG-STLVTPNPNNGFLACGPLYAYRCGH 140
Qy 135 SVFSSGICARVDASFQPGQSLAPQRCPTVMDVIVLDGNSNIYPWSEVQTFRLRLVGLK 194
Db 141 LHYTTGICSDVSPFTQVNSIAP-VQECSTQLDIVILDGNSNIYPWSDVTFAPLNDLLE 199
Qy 195 LFTDPEIQIQLVQYSGSPVHNSLGDFTKEEVVRAAKNLSRRREGRETCTAQAIMVACT 254
Db 200 MDIGPKQTQVGIQYQGENVTHEFNLNKYSSTEEVLVAAKKIVQRGGRQCTWTALGIDTARK 259
Qy 255 EGFSSQSHGGRPEAARLLVVTDGSHSGEELPAALKACEAGRVTRYIAVLGHYLRQRD 314
```

Db 260 EAFTEARGRGVKKVMVITVDGESHNDHRLKVKVIQCEDENIQRFSAIILGSYNRGNLS 319  
Qy 315 PSSFLREIRTIASDPDERFFNVNDEAALTDIVDALGDRIFGLEGSHAENESSFGLEMSQ 374  
Db 320 TEKFEVEIKSIASEPTEKHFNVDELALVTIVKTGLGERIPALEATADQSAASEMEMSQ 379  
Qy 375 IGFSHRLKDGILSGMVGAYDMGCVLWLEGGHRLFPFRMALEDEFFPALQNHAAVILGYS 434  
Db 380 TGPSAHYSQDMVMLGAGVAYDMGVTVMVQKASQIIIPRNTTFNVESTKKNPLASYLYGT 439  
Qy 435 VSSMLLRGRLFLSGAPFRHRGKVIAPOLKDKGAVRVAQSQLOEQIGSYFGSELCPDL 494  
Db 440 VNSATASGDVLYIAGQPRYNHTQVLIYRM-EDGNIKILTSGEQIGSYFGSILITTD 498  
Qy 495 TDRGTTDVLVAAPMFLGPQNKETGRVYVYLVGQOQLLTLOGTLP----- 541  
Db 499 IDKDSNTDILLVGAPVMVNGTEKEQGVVYVAL-NQTRFEYQMSLEPIKQTCSSRQHNS 557  
Qy 542 -----EPPQDARFGFANGALPDLNODGFADVAVGAPLEDHOGNALYLYHGTQSGVRPH 595  
Db 558 CTTENKNEPCGAREGTAIAVKDNLNODGFNDIVIGAPLEDHGGAVIYHSGGKTIKEY 617  
Qy 596 AQRIAAASMPHALSYFGRSVDRGLDLDGDDLVDAVGAQGAAILSSRPVHLTPSLEVT 655  
Db 618 AQRIPSGCGDKTLKFFGSGIHGMDLNGDGLTDVTIGGLGAALFWSRDVAVVKVWNFE 677  
Qy 656 POAISVQDRCRRGQEAVALCFOVTSRTFGRWDHOFYMRFTASLDEWTAGARAAF 715  
Db 678 PNKVINIKKNCHMEGKETVCINATVCFDVKLKSKEDTIYEADLQYRVTLDSLRQISRSFF 737  
Qy 716 DSGQRSLPRRLRLSVGNVTCOEQLHFHVLDTSDYLRPVVALTVFALDNTTKP--GPVLNE 773  
Db 738 SGTQERKQVR--NITVRKSECTKHSFYML--HDFQDSVRITLDF---NLTDPENGPVLDD 790  
Qy 774 GSPTSIOKLVPFKDCGPDNCEVTDVLQVNMDIRGSRKAPFVVRGGRKVLVSTLTENR 833  
Db 791 SLPNSVHEYIIFAKDCGKKEKISDLSLV---ATTEKDLIIVRSQNDKFNVSLTVKNT 846  
Qy 834 KENAYNTSLSIIFSRNLHSLTPQRESPIKVECAAPSAHARLCSVGHVPVQTGAKVTL 893  
Db 847 KDSAYNTRTIVHYSPLVFSGI---EAIQKDCS--ESNHNITCKVGYFFLRRGEMVTFK 900  
Qy 894 LEPEFCSLLSQVFGKLTASSDSLERNGTLOENTAQTSAYIOYEPHLLFSSESTLHRYE 953  
Db 901 ILFOFNTSYLMENVTIYLSATSDSEPPETLSDNVNVNISIPVKYEVGLQFYSSASEYHIS 960  
Qy 954 VHPYGTLP-----VG-----PGPEFKTLRTNNAACIVQNLTEPPGP 990  
Db 961 IAAKETVPEVINSTEDIGNEINI FYLIRKSGSFPMPELKLIS-----FPNMTSNGYP 1013  
Qy 991 FVHPEELQHTNRLN-----GSNTQCO-----VVRCHLG 1018  
Db 1014 VLYPTGLSSSENANCRPHI PEDPPFSINGKMTTSTDHLKRGITLDCNTCKFAITCNLT 1073  
Qy 1019 QLAGTEVSVGLRLVHNHFRRAKFKSLTVVSTFELCTEGSVLQLTEASRWSESLELV 1078  
Db 1074 S-SDISQVNVSL--ILWKPTFKSYFSSINLTIRGELSENAS-LVLSSNQKRELAIQI 1129  
Qy 1079 VQT-RPILISLWILIGSVLGLLALLVFLCWLKGFPAHKKIPEEKREE 1128  
Db 1130 SKDGLPGRVPLWILLAFAGLLMLLILALWKIGFF---KRPDKKQNEK 1177

Search completed: April 6, 2005, 12:53:09

Job time : 148.389 secs

**This Page Blank (uspto)**



Db	578	PRQYFGQALSGQDLTQDGLMDLAVGARGQVLLKSLPVLKVGVMRFSPEVAKAYVR	633
Qy	666	C-----RRRQGEAVCLTAALCFQVTSRTGRWDHOFVMTASLSDWTAGAAAPDGS	718
Db	638	CWEEKPSALEAGDATVCLTIQ-----KSLDQLGDIQSSVRFDLALDPRITSAIFNET	692
Qy	719	GORLSRRRLSLVGNVTCBQLHFHVLDTSDYLRVALTVTFALDNTTKPG-----PVLN	772
Db	693	KNPTLFRKNTGLG-IHCETLKLPLDCVEDVVSPIILHLNFSLVREPIPSQNLRLPULA	751
Qy	773	EGSPTSIQKLVFSPKDCGPDNECVTLVLQVMDIRGSRKAPFVVVRGGRKVLSTLLEN	832
Db	752	VGSQDLFTASLPFEKNCQDGLCEGD--LGVTLSPSGLQT--LTVGSSLELNVITVYN	806
Qy	833	RKENAVNTSLSIIFSRNLHLASLTQRES-----LKVSC-AAPSAHARL---CSVGHVP	883
Db	807	AGEDSYGTVVVSUYIPAGLSHRRVGAQKQPHOSARLACETVPTDEGLRSRCSVNHI	866
Qy	884	FOTGAKVTFLLEFEFSCSLLSQVFKLTAGSDSLERNGTQJQNTAQTSAYIQEPHLLF	943
Db	867	FHEGSNGTIVTFVDSYKATLGRM-LMRASSENKASSKATFQLELPKVAVTMI	925
Qy	944	S---SESTL-----HRYEVHPVGTLPVGPGBEFKTTLTNNASCIVQMLT	985
Db	926	SROESTKYFNFAISDEKQKAEHRYRVNLSQRDLAISINFWVPVLLNGVA-VWDVVM	984
Qy	986	EPGP-----PVPHEELOHTNR---LNGSNTQCVVRCHLQGLAKGTEVSGLLR	1033
Db	985	EAPSQSLPCVSRKPKQSHSDFLTQISRPMLDCSIADCLQRCDFVPSFQVELDFTLKG	1044
Qy	1033	LVNEFFRAKPKSLTVSVTFELGTEGSLVQLTEASRWSLSLE-VVQTRPILISLWIL	1091
Db	1045	NLSFGVRETQKVLVSVAEITFTSVYSQLPQEAQFMAQEMVLEEDVYNAIPII	1104
Qy	1092	IGSVLGLLLALLVFLMKLQFF-AHKKIPBEKREE	1128
Db	1105	MGSSVGAALLALITATLYKLGFRRKRVKEMLEKPED	1142
RESULT 14			
US-08-943-363-2			
Sequence 2, Application US/08943363			
Patent No. 5837478			
GENERAL INFORMATION:			
APPLICANT: Gallatin, W. Michael			
TITLE OF INVENTION: Van der Vieren, Monica			
NUMBER OF SEQUENCES: 114			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun			
STREET: 233 South Wacker Drive, 6300 Sear Tower			
CITY: Chicago			
STATE: Illinois			
COUNTRY: United States			
ZIP: 60606-6402			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: Patent In Release #1.0, Version #1.25			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/943,363			
FILING DATE:			
CLASSIFICATION: 530			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 08/173,497			
FILING DATE: 23-DEC-1993			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 08/286,889			
FILING DATE: 5-AUG-1994			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 08/362,652			

Db	578	PRQYFGQALSGQDLTQDGLMDLAVGARGQVLLKSLPVLKVGVMRFSPEVAKAYVR	633
Qy	666	C-----RRRQGEAVCLTAALCFQVTSRTGRWDHOFVMTASLSDWTAGAAAPDGS	718
Db	638	CWEEKPSALEAGDATVCLTIQ-----KSLDQLGDIQSSVRFDLALDPRITSAIFNET	692
Qy	719	GORLSRRRLSLSVGNVTCBQLHFHVLDTSDYLREVALTVTFALDNTTKPG-----PVLN	772
Db	693	KNPTLFRKNTGLG-IHCETLKLPLDCVEDVVSPIILHLNFSLVREPIPSQNLRLPULA	751
Qy	773	EGSPTSIQKLVFSPKDCGPDNECVTLVLQVMDIRGSRKAPFVVVRGGRKVLSTLLEN	832
Db	752	VGSQDLFTASLPFEKNCGQDGLCEGD--LGVTLSPSGLQT--LTVGSSLELNVITVYN	806
Qy	833	RKENAVNTSLSIIFSRNLHLASLTQRES-----LKVSC-AAPSAHARL---CSVGHVP	883
Db	807	AGEDSYGTVVVSUYIPAGLSHRRVSGAQKQPHOSARLACETVPTDEGLRSRCSVNHI	866
Qy	884	FOTGAKVTFLLEFEFSCSLLSQVFKLTAGSDSLERNGTQJQNTAQTSAYIQEPHLLF	943
Db	867	FHEGSNGTIVTFVDSYKATLGRM-LMRASSENKASSKATFQLELPVKVAVYTM	925
Qy	944	S--SESTL-----HRYEVHPVGTLPVGPGBEFKTKTLTNASCIVQMLT	985
Db	926	SROESTKYFNFAISDEKQKAEHRYRVNLSQRDLAISINFWVPVLLNGVA-VWDVVM	984
Qy	986	EPGP-----PPVHPEELOHTNR---LNGSNTQCVVRCHLQGLAKGTEVSVGLLR	1033
Db	985	EAPSQSLPCVSRKPKQSHSDFLTQISRPMDCSADCLQRCDFVPSVQOEELDTLKG	1044
Qy	1033	LVNEHFRRAKPSLTVVSTFELGTEGSLVQLTEASRWSLSLE-VVQTRPILISLWIL	1091
Db	1045	NLSFGVRETLOKVLVSVAEITFTSVYSQLPQEAQFMAQEMVLEEDVYNAILI	1104
Qy	1092	IGSVLGLLLALLVFLMKLQFF-AHKKIPBEKREE	1128
Db	1105	MGSSVGAALLALITATLYKLGFRRKHKEMLEDKPED	1142
RESULT 14			
US-08-943-363-2			
Sequence 2, Application US/08943363			
Patent No. 5837478			
GENERAL INFORMATION:			
APPLICANT: Gallatin, W. Michael			
TITLE OF INVENTION: Van der Vieren, Monica			
NUMBER OF SEQUENCES: 114			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun			
STREET: 233 South Wacker Drive, 6300 Sear Tower			
CITY: Chicago			
STATE: Illinois			
COUNTRY: United States			
ZIP: 60606-6402			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: Patent In Release #1.0, Version #1.25			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/943,363			
FILING DATE:			
CLASSIFICATION: 530			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 08/173,497			
FILING DATE: 23-DEC-1993			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 08/286,889			
FILING DATE: 5-AUG-1994			
PRIOR APPLICATION DATA:			
APPLICATION NUMBER: US 08/362,652			



Qy	773	EGSP	TSIQKLVPRFSGKDPNCEVTDVLVQNMDDIRGSRKAPVWVGGRKVLVSTTLN	833
Db	752	VGSQ	DLFTASLPFFKNCQDGLCEGD--LGVTLFSGQLQ---LTVGSSLELNVIIVTVM	806
Qy	833	RKENAY	NTSLSGIIFSRNLHLASLTPORESF-----IKVEC-AAPSAHARL-----CSVGHVP	883
Db	807	AGE	SYGTVWSLYYPAGLSHRVSGCAQKQPHQSALRLACETVPTDEGLRSRCSVNHDI	866
Qy	884	FQTGA	KVTFLLFEPFSCSSLLSQVFGKLTASSDSLRNGTLOBNTAQTSAIQYEPHLLF	943
Db	867	FHEG	SGTFTVTFDVSYKATLDGRM-LMRASASSENKASSKATFOLELPVKYAVYTWI	925
Qy	944	S--SESTL	-----HRYEVPIPYGTLPVGPQPEFKTTLRTTNASCIQVNLIT	985
Db	926	SQREEST	KYFNFATSDKKMKBAHRVYNVNLQDRLAISINFWPVLVLLNGVA-VMDVVM	984
Qy	986	BPG	-----PPVHPELOHTNR---LNGSNTQCQVVRCHLGOLAKGTESVGLLR	103
Db	985	EAP	SQSLPCVSRKPPQHSDFLTQISRSPMDCSIADCLQFRCDVPSFSVQELDFTLKG	104
Qy	1033	LVHNE	FFRRAKSLTVVSTFELGTESGVQLQTEASRMSESLLE-VVQTRPILISLWIL	109
Db	1045	NLSFG	WRETLQKKVLVSVVAEITFDTSVYSQLPGQEAQFMRQWEMVLEDEYVNAIPII	110
Qy	1092	IGSV	LGGLLALLVFLKWLKGF-AHKKPIPEEKREE	1128
Db	1105	MGSSV	GALLLALITATLYKLGFPRHYKEMLEDKPED	1142
RESULT 15				
US-09-193-043-2				
; Sequence 2, Application US/09193043				
; Patent No. 6251395				
; GENERAL INFORMATION:				
; APPLICANT: Gallatin, Michael W.				
; APPLICANT: Van der Vliet, Monica				
; TITLE OF INVENTION: NO. 6251395el Human 2				
; FILE REFERENCE: 27866/35004				
; CURRENT APPLICATION NUMBER: US/09/193,043				
; EARLIER FILING DATE: 1998-11-16				
; EARLIER APPLICATION NUMBER: 08/173,497				
; EARLIER FILING DATE: 1993-12-23				
; EARLIER APPLICATION NUMBER: 08/286,889				
; EARLIER FILING DATE: 1994-08-05				
; EARLIER APPLICATION NUMBER: 08/362,652				
; EARLIER FILING DATE: 1994-12-21				
; EARLIER APPLICATION NUMBER: 08/943,363				
; EARLIER FILING DATE: 1997-10-03				
; NUMBER OF SEQ ID NOS: 114				
; SOFTWARE: PatentIn Ver. 2.0				
; SEQ ID NO 2				
; LENGTH: 1161				
; TYPE: PRT				
; ORGANISM: Homo sapiens				
US-09-193-043-2				
Query Match 19.0%; Score 1126; DB 3; Length 1161;				
Best Local Similarity 29.5%; Pred. No. 7 5e-99;				
Matches 348; Conservative 203; Mismatches 503; Indels 124; Gaps 41				
Qy	23	FNLDE	HHRLFPFGPPEAFEGYSVLQHVGGGQRMWLVGAPWDGSGDRRGDVTRCPVGGAH	82
Db	17	FNL	VEEPTIFQ-EDAGGFGGQSVVQF--GGSR-LVVGAPLEVVAAANTGLRYDC--AAA	69
Qy	83	NAPCA	KHGLDYQLGNSSHPAVNMHLGMSLLETTDGGGFACAPLWSRACGSSVFSSGIC	142
Db	70	TGMCQ	PIPL-----HIRPEAVNNSLGLTLAASTNGSRLLACGPTLHRVCGENSYSKGSC	123
Qy	143	ARVDA	SFQPGSLAPTAQRCP-TYMDVVIVLDGNSNI--YPWSEVQTFRLRLVGLKLFIDP	199
Db	124	LLLSGR	WIIITQVTDPECPHQEMDIFVLIDGSGIDQNDNMKGFFVQAVMGQ--FSG	181
Qy	200	EOIGV	GLVOYGESPVHWSLGDFTKEBVRRAKNLSRREGRETAKTAQALMWACTEGFSG	259

```
Db      182  TDTLFAIMQSNLKIHTFTQFRTSPSQSLVDPIVQLKGL-TFTATGILTVVTQLFHH 240
Qy      260  SHGGRPEAARLLVVVTGESH-DGEELPAALKACEAGRVTRYGIATVGHVLRQRDPSSF 318
Db      241  KNGARKSAKKILVITDQKIDPLEYSDVIPQAEKAGIIRYAIGV-GHAF---QGFTA- 295
Qy      319  LREIRTIASPPDERFFNFVDEAALTDIVDALGDRIIFGLEGSHAENESSFGLEMSQIGFS 378
Db      296  RQELNTISSAPPQDHVFKVDNFALGSIQKOLQEKIYAVEGTQSRASSSQHEMSQEGFS 355
Qy      379  THRUKDGLFQMGVAYDWGSGVLMLEGGHRLFPPRMALEDEFPFALQNH-----AY 430
Db      356  TALTMGDLFLGAVGSFSW-----SGGAFLYPPNMS-----PTFINMSQENVDMRDSY 402
Qy      431  LGSVSSMLLRGRRLLFSGAPRRHRGKVIAT-QLAKKGAVRVAQSLQEQIGSYEGSE 489
Db      403  LGSYTELALWGVQNLVL-GAPRYQHTGKAVIFTQVSRQW--RKKAETGTGIGSYFGAS 459
Qy      490  LCPLDTRDGTDTVLLVAAPMFLGPQNKETGRVYVYL--GQOSLLTLQGTLOPEPPQD- 546
Db      460  LCSVDVDSGSDTLLILGAPHY--EOTRGGQVSVCPLPGRQVQWQCDVLRGEQGHWP 517
Qy      547  ARFCFANGALPDLNQDGFADVAGAPLEDGHQGALYLYHG-TQSGVRPHPAQRIAAAMP 605
Db      518  GRFGAALTVLGDVNEDKLIDVAIGAPGEQENRGAVYLFHGASESGISPSHSQRIASSQLS 577
Qy      606  HALSYFGRSVDRLDLDGDDLDVAVGAGQAAILSSRPVHLTPSLEVTPOAISVVQRD 665
Db      578  PRLQYFOGALSQGGQDLTQDGLMDLAVGARGQVLLRLSLPVLKGVAMRFSPEVAKAVYR 637
Qy      666  C-----RRRQGBAVALTAALCFQVTSRTPGRWHDQFYMRFTASLDEWTAGARAAFDGS 718
Db      638  CWEKPSSALEAGDATVCLTIQ-----KSSLDQLGDIQSSVRFDLALDPGLTSAIFNET 692
Qy      719  GQRLSPRLRLSVGNVTCQLHFHVL-D-TSYLRPVALTVTFALDNTTKPG-----PVLN 772
Db      693  KNPTLTRRKTGLG- IHCETLKLPLDCVEDVWSPIILHLNFSLVREPIPSQNLRPVLA 751
Qy      773  EGSPTSQKLVPPFSKDCPDNECVTDVLQVNMDIRGSRKAPFVVRGRRKVLVSTILEN 832
Db      752  VGSQDLFTASLPFFBKNCQDGLCEGD--LGVTLFSFGLT---LTVGSSLELNVIVTWN 806
Qy      833  RKENAYNTLSIIFSRNLHLASLTPQRESP----IKVEC-AAPSAHARL-----CSVGHPV 883
Db      807  AGEDSYGTVVSLYYPAGLSHRVSVGAQKQPHQSALRLACETVPTDEGLRSSRCVNHPI 866
Qy      884  PQTGAKVTFLLBFFSCSSLLSQVFGKLTASSDSLERNGLQENTAOQTSAYIQYEPHLLF 943
Db      867  PHEGSGTFTVFDVSVYKATLGDRM-LMRASASSENKASSKATFQLELPVKYAVVTMI 925
Qy      944  S--SESL-----HRYEVHPYGTLPVGPPEPKTTLRTNNASCIVQNL 985
Db      926  SRQESTKYFNATSDKKMKAEHRYRVNLSQRLAISINFWPVLLNGVA-VMDVVM 984
Qy      986  BPPG-----PPVHPEELQTNR---LNGNTQOVVRCHLQOLAKGTEVSVGLLR 1032
Db      985  EAPQSILPCVSRKPPQHSDFLTQISRSPMLDCSIADCLQRCDDVPSPVOEELDFTLKG 1044
Qy      1033  LVHNEFFRRAKFKSLTVVSTPELGTGEGSVLQLTEAGRWSSLEL-VVQTRPILISLWIL 1091
Db      1045  NLSFGWVRETLQKKVLVSVVAEITFDTSVYSQLPQEAQMRQAQMEMVLEEDVYNAIPII 1104
Qy      1092  IGSVLGGILLALLVCLWKLGF- AHKKIPEEKREE 1128
Db      1105  MGSSVGALLLALITATLYKLIGFKRHYKEMLEDKPED 1142
```

Search completed: April 6, 2005, 12:24:46  
Job time : 72.3117 secs

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6189
; LENGTH: 1181
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6189

Query Match          29.2%; Score 1733.5; DB 4; Length 1181;
Best Local Similarity 33.7%; Pred. No. 1.6e-157;
Matches 410; Conservative 217; Mismatches 431; Indels 159; Gaps 32;

Qy 11 LPLVFLTGL-----CSPFNLDHHPRLFPGPPEAEFGYSLQHVGGQGRWMLVGAPWD 63
Db 11 LPLLLVLAISQILNCLAYNVGLPEAKIFSGPSEQFGYAVQVQFINPKGNWLLVGSFWS 70

Qy 64 GPSGDRGDVYRCPVGGAHNAPCAKAGHL-GDYQLGNSHPAVNMHLGMSLLETDCGGFM 122
Db 71 GFPENRMGDVYKCPV-DLSTATCEKLNQTSIPNVTMKTNNLSGLILTRNMGTCGFL 129

Qy 123 ACAPLWSRACSSVFSGGICARVDASPOQSLAPTAQRCPTYMDVIVLDGNSIYPWS 182
Db 130 TCGPLMAQQCGNQYTTTGVCSIDIPDFQLSASFSPATQPCPSLDIVVWVCDENSIYPWD 189

Qy 183 EVOTFLRLVGLKFLIDPEIQVGLVQGESPVHWSLGDPRTKBEVVRRAAKNLSRRGRE 242
Db 190 AVKNFLEKFOVGLDIGPTKTQVGLIQYANNPRVFNLTYYTKKEMIVATSQTSQYGGDL 249

Qy 243 TKTAQAIMVACTEFGSQSHGGRPEARLLVVVTVDGSHDGBELPAALKACBAGRVTYGI 302
Db 250 TMTFCAIQARKYAYSAAAGRRSATKVMVVVTDGSHDGSMLKAVIDQCNDHNLRFGI 309

Qy 303 AVLGHYLRQRDPSSFLUREITIASDPPERFFNVNTDEAALTDIVDALGDRIFGLESHA 362
Db 310 AVLGYLRNALDTKNLKEIKAIASIPTEYFFNVDSDEAALKEAGTLGEQIFSIETG-V 368

Qy 363 ENESFGLMSQIGFST--HRLKDGILFGMGAYDGGSVLW-LEGGHRLPPRMALDE 419
Db 369 QGGNFQWMSQVGSADYSSQNDILMLGAVGAFGWSGTIVQKTSHGHLIFP-----KQA 423

Qy 420 PPPALQ--NHAAYLGYSVSSMLLGGRRFLPSGAPRFRHRGKVIAFQKKDGAVRVQSL 477
Db 424 FDQILQDRNHSSYLGYSVAA-ISTGESHFVAGAPRANYTQGIVLVSVNENGNTIVQAH 482

Qy 478 QGEQISYFGSELCPDLTDRTGGTDLVLAAPMFLGPONKETGRVYV-----LVGQSL 532
Db 483 RGDQIGSYFGSVLCSDVDKDTITDVLVAGAPMTMSDLKKEGRRVYLFITKKGILGQHF 542

Qy 533 LTLOGTLOPEPPODARFGFANGALPDNLQDGFADVAVGAPLEDHOGALYLYHGTQSGVR 592
Db 543 --LEG---PEGIENTRFGSAIAALSDINMGDFNDVVGSPLENQSGAVIYNGHQGTIR 597

Qy 593 PHPAQRIAAA--SMPHALSYFGSRVDGRDLDDGDLVDVAVGAQCAAILLSRRPIVHLTP 650
Db 598 TKYSQKILGSDGAFRSHLOYFGRSLDYGDLNGDSITDVSIGAFQVQVQLWSQSIADVAI 657

Qy 651 SLEVTPOAISVVQRDCRRRGOEA VCLTAALCFQVTSRTPGGRDHFQYNNRFTASLDWTFAG 710
Db 658 EASFTPEKITLVNNAQ-----IILKCFSAKFR--PTKQNNQVAIVNITLDA----- 704

Qy 711 ARAAFDGSQRLSPRL-----RLSVGNV-----TCEQLHFHVLDTSDYLRPVALTVT 758
Db 705 -----DGFSSVTRSGFLPKENNERCLQKNVMVNAQSCPEHIIYIQEPSDVVNSLDLRVD 759

Qy 759 PALDNTTKPG--PVLNESPSTSIQKLVFPFSKDCPDNECVTDLVLQVNMDIRGSKAPFV 816
Db 760 ISLEN--PGTSPALAEVSETAKVFSIFPHKDCGEDGLCISDLVDLVR-QIPAAQEQPFI 815

Qy 817 VGRGRRKVLVSTTLNKRKENAYNTSLSIIFSRNLHLASLTPORESPI---KVEC-AAPSA 872
Db 816 VSNQKRLTFTVLNKRKESAYNTGI VVDFSENLFASFSS---LPVDGTETVTCQVAASQ 871

Qy 873 HARLCSGVHPVFOGAKVTFLEPEFSCSSLLSQVFGKLTASSDSLEBNGTLQENTAGTS 932
Db 932
```

```
872 KSVACDVGPALKREQVTTINFDNQLQNLQNASLSFQALSSESQENKA--DNLVNLK 929
933 AYIQYEPHLLFSSESTLHYEVHPYGTLP-----VGPGEFKTTLTNNASCIQV---- 982
930 IPLLDAEIHILTRSTNINFYEISSDGNVPSIVHSFEDVGPKFIFSLKVTGTSVFVSMATV 989
983 -----NLTEPPGP-----PVHP-----BELQHTNRLN 1004
990 IIHPIQYTKENPLMYLTGVQTKAGDISCNADINPLKIGQTSSTSVFSKSNFHTKELN 1049
1005 GSNTQCVVRCHLQGLAKGTEVSGLLRVLVHNEFFRRAKFKSLTVVSTFELGTBEGSVLQ 1064
1050 CRTASCNVTCWLDKDVHMKGEYFNVVTRTWNGTFASSTFTQVLTAAAEINTVNPBIY- 1108
1065 LTEASRWSESLLEVVQTRPILISLWI-----LIGSVLGGLLLLALLVFCWLK 1111
1109 -----VIEDNTVTIPLIMKPDKAEVPTGVIIGSIAGIILLALLLVAIWLK 1155
1112 LGFFPAHK-----KIPER 1123
1156 LGFFKRKYEKMTKNPDE 1172

RESULT 6
US-09-949-016-10747
; Sequence 10747, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10747
; LENGTH: 1195
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10747

Query Match          29.2%; Score 1733.5; DB 4; Length 1195;
Best Local Similarity 33.7%; Pred. No. 1.6e-157;
Matches 410; Conservative 217; Mismatches 431; Indels 159; Gaps 32;

Qy 11 LPLVFLTGL-----CSPFNLDHHPRLFPGPPEAEFGYSLQHVGGQGRWMLVGAPWD 63
Db 25 LPLLLVLAISQILNCLAYNVGLPEAKIFSGPSEQFGYAVQVQFINPKGNWLLVGSFWS 84

Qy 64 GPSGDRGDVYRCPVGGAHNAPCAKAGHL-GDYQLGNSHPAVNMHLGMSLLETDCGGFM 122
Db 85 GFPENRMGDVYKCPV-DLSTATCEKLNQTSIPNVTMKTNNLSGLILTRNMGTCGFL 143

Qy 123 ACAPLWSRACSSVFSGGICARVDASPOQSLAPTAQRCPTYMDVIVLDGNSIYPWS 182
Db 144 TCGPLMAQQCGNQYTTTGVCSIDIPDFQLSASFSPATQPCPSLDIVVWVCDENSIYPWD 203

Qy 183 EVOTFLRLVGLKFLIDPEIQVGLVQGESPVHWSLGDPRTKBEVVRRAAKNLSRRGRE 242
Db 204 AVKNFLEKFOVGLDIGPTKTQVGLIQYANNPRVFNLTYYTKKEMIVATSQTSQYGGDL 263

Qy 243 TKTAQAIMVACTEFGSQSHGGRPEARLLVVVTVDGSHDGBELPAALKACBAGRVTYGI 302
Db 264 TMTFCAIQARKYAYSAAAGRRSATKVMVVVTDGSHDGSMLKAVIDQCNDHNLRFGI 323

Qy 303 AVLGHYLRQRDPSSFLUREITIASDPPERFFNVNTDEAALTDIVDALGDRIFGLESHA 362
Db 362
```

Db 324 AVGLYNRLALDTKNLIKATASIPTEYFFNVSDAALKEKAGTLGHEQIFSIET-V 382  
Qy 363 ENSSFGLEMSQIGFST--HRLKDGILFGVGVADMGGSVLM-LEGGRHLFPPRMALEDE 419  
Db 383 QGGDNFQWMSQVGSADYSQNDILMLGAVGAFMGSGTIVQKTSHGHLIFP-----KQA 437  
Qy 420 FPPALQ--NHAAYLGYSVSMLLRGRRLLFSGAPRHRGKVIQFOLKXDGAVRVAQSL 477  
Db 438 FDQILQDNHSSVLYGSVAA-ISTGESHTFVAGAPRANYTGOIVLYSVNENGNITVIAQ 496  
Qy 478 QGEQIGSYFQSELCPLDTRDGTVDVLLVAAPMFLGPONKETGRVYVY-----LVGQOQL 532  
Db 497 RGDQIGSYFQSVLCSDVDKDTITDVLVAGAPMYSDLKKEBGRVYLTIKGILGQHOF 556  
Qy 533 LTLOGLTQPEPPQDARFGPANGALPDINODGFADVAGAPLEDHOGALVLYHCTQSGVR 592  
Db 557 --LEG---PGIENTRFGSAIALUSDINMGDFNVIIVGSPLENQSGAVVIYNGHQGTIR 611  
Qy 593 PHPAQRIAAA--SMPHALSFGRSVGRQLDLXGDDLVDAVAGAGAAILLSSRPVHLTP 650  
Db 612 TKYSQKILGSDGAPRSHLYFGRSLDGYDGLNGDSITDVISGAFQGVVQLMSQSIADVAI 671  
Qy 651 SLEVTQAIISVQDRRRRQGEAVCLTAALCFQVTSRTPGKWDHQFYMRFTASLDEWTAG 710  
Db 672 EASPTPEKITLVNKAQ-----IILKLCFSAPFR-PTKQNNQVAIVYVNTILDA--- 718  
Qy 711 ARAAFDGGGQRLSPRL-----RLSVGNV-----TCEQLHFHVLTSDYLRVALTVT 758  
Db 719 -----DGFSSRVTSRGLFKENNERCLQKNVMVNOAQCPBHIYIQBPDSVMSLDLRVD 773  
Qy 759 FALDNTTKPG--PVLNBSGPTSIOKLVPFSDKQPDNECVTDVLQVNMIDIRGRKAPFV 816  
Db 774 ISLEN---PGTSPALEAYSETAKVFSIPFKDCGEDGLCTSDLDVDR-QIPAAQEQPFI 829  
Qy 817 VRGGRKVLVSTTLNENKAYNTSLSIIFSRNLHLASLTPQRESPI---KVEC-AAPSA 872  
Db 830 VSNQKRLTFSVTLNKRASAYNTGIIVDFSENLPFASFS-----LPVDGTEVTCQVAASQ 885  
Qy 873 HARLCSVGHVPFGAKVPTLLEPFSCSLLSQVFGKLTASSDSLERNGTLOENTAQTS 932  
Db 886 KSVACDVGYPALKREQQVTTINFDFNLQNLQNASISFQALSSEBENKA--DNLVNLK 943  
Qy 933 AYIOYEPHLLFSSBSTLHRYEVHPYGTLP-----VGPGEFKTLRTNNASCIVQ--- 982  
Db 944 IPLYDAEHLTRSTNINFVISDGNVPSIVHSFEDVGPKFIPSLKVTTCSPVSNATV 1003  
Qy 983 -----NLTEPPGP-----PVHP-----FELQHTNRLN 1004  
Db 1004 IHIPOYTKBNPLMYLTGVQTDKAGDISCNADINPLKIGQTSSSSVSFKSENFRHTKELN 1063  
Qy 1005 GSNTQCVVRCHLQLAGTEVSGLLRLVHNEPFRAPKSLTVGTFELGTBEGSVLQ 1064  
Db 1064 CRTASCNVTCWLKXDVHMKGBYFNVVTRIMNGTFASSTFQTQVLTAAEINTYNEIY- 1122  
Qy 1065 LTESRWSESLEVVQTRPILISLWI-----LIGSVLGGLLIALLVFLCWK 1111  
Db 1123 -----VIEDNTVTIPLMKPDEKAEVPTGVIIGSIAGILLALLVALVILWK 1169  
Qy 1112 LGFFPAHK-----KIPEE 1123  
Db 1170 LGFFPKRYEKMKNPDE 1186

RESULT 7  
US-09-532-310B-6  
Sequence 6, Application US/09532310B  
Patent No. 6596276  
GENERAL INFORMATION:  
APPLICANT: Senger, Donald R  
Detmar, Michael  
Claffey, Kevin P  
TITLE OF INVENTION: Method for inhibiting tumor

angiogenesis in a living subject  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David Prashker, Esq.  
STREET: P.O. Box 5387  
CITY: Magnolia  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 01910  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage  
COMPUTER: Dell PC  
OPERATING SYSTEM: MS DOS  
SOFTWARE: Microsoft Word version 97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/532,310B  
FILING DATE: 22-Mar-2000  
CLASSIFICATION: Unknown  
ATTORNEY/AGENT INFORMATION:  
NAME: David Prashker, Esq.  
REGISTRATION NUMBER: 29,693  
REFERENCE/DOCKET NUMBER: BIS-036  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (978) 525-3794  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1183 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-532-310B-6

Query Match 28.7%; Score 1704.5; DB 4; Length 1183;  
Best Local Similarity 33.0%; Pred. No. 9,9e-155;  
Matches 404; Conservative 227; Mismatches 435; Indels 157; Gaps 33;  
Qy 11 LPLVFLTGL-----CSPENLDEHHPRLPPGPEABERGYSVLQHVGGQGRWMLVGAPWD 63  
Db 11 LPLLLVLAUSQGLINCLAYNGLPEAKIPSGSSQFGYAVQOQINPKGNWLLVSGSPWS 70  
Qy 64 GPSGDRRGDYVRCPVGAHNAAPCAKQHL-GDYQLGNSSHPAVNMHLGMSLLETDDGGPMD 122  
Db 71 GPPENRMGDVYKCPV-DLSTATCEKLNLTQSTSI PNVTETKTNMSLGLILTRNMTGGFL 129  
Qy 123 ACAPLMSRACSSVFSGICARVDASFQPOGSLAPTAQRCPTVMDVIVJDGNSIYFWS 182  
Db 130 TCGPLWAQCGNQYTTGVGSDISDPDFXKSASFSPATXPCPSLIDVVVVCDSNSIYPWD 189  
Qy 183 EVQTELRRLVGLKLPIDPEQIQVGLVOYGESPVHEMSIGDPRTKEEVVRAAKNLSRRREGRE 242  
Db 190 AVKNFLEKFGQIDIGTKTQVGLIQYANNPRVFNLTNTYKTEEMIVATQSQSYGGDL 249  
Qy 243 TKTAQAIMVACTSGFSQSHGRPEEARLLVVVTDGSHDGEELPAALKACEAGRVTRYGI 302  
Db 250 TNTFGAIQYARKAYSAAGSRSSATKVMVVTVDGSHDGSMLKAVIDQCNDHNDILAFGI 309  
Qy 303 AVLGHYLRORDPS-SFLRIRTIASDPDRFPFNVTDEAALTDIVDALGDRIFGLBGSH 361  
Db 310 AVLYLNRNALDTKNMLIKEIKAISIPTEYFFNVSDAALKEKAGTLGHEQIFSIET- 368  
Qy 362 AENESSFGLMSQIGFST--HRLKDGILFGVGVADMGGSVLM-LEGGRHLFPPRM--AL 416  
Db 369 VXGDNFQWMSQVGSADYSQNDILMLGAVGAFMGSGTIVQKTSHGHLIFPQAQDOI 428  
Qy 417 EDEFPFALQNHAAAYLGYSVSMLLRGRRLLFSGAPRHRGKVIQFOLKXDGAVRVAQS 476  
Db 429 QD-----RNHSSVLYGSVAA-ISTGESHTFVAGAPRANYTGOIVLYSVNENGNITVIA 481  
Qy 477 LQGEQIGSYFQSELCPLDTRDGTVDVLLVAAPMFLGPONKETGRVYVY-----LVGQOQL 531  
Db 482 HRGDQIGSYFQSVLCSDVDKDTITDVLVAGAPMYSDLKKEBGRVYLTIKGILGQHOF 541

532	Qy	LLTILQTLQEPBPPODARFGFAMGALPDLINODGFADVAVGAPLEBDHGOGALYLHYGTOSGV	591
542	Db	F--LEG---PEGIENTFGSAIATSLINMDGNFNDVITGSPLEQNQNSGAVIYNGHOGTI	596
592	Qy	RHPAPQRIAAA--SMPHALSIFYGRSDVGRDLDOGDLDVAVAGQAAGAAIILLSRPIVHLT	649
597	Db	RTKYSQILGSDGAFRSHLQYFGRSLDYGDLNGDSITDVSGIAGFQVVQLMSQSIADVA	656
650	Qy	PSLEVTPOALISVVORDCRRRGOEAVCLUTAAALCFQVTSRTGRWDHGHQYMRFTASLDSEWTA	709
657	Db	IEASFTEPKITLVNKNQAQ-----IILKLCFSAKFR-PTKQNNQVAIVYNTILDA---	704
710	Qy	GARAAAPGSGORLSPRRL-----RLSVGNV-----TCBQLGHEVLDTSDYLPRV--ALT	756
705	Db	-----DGFSSRVTSRGLPKENNERCLQKNVYVNAQSCPEHIYIQEPEDVNVNSLDDL	758
757	Qy	VTFALDNTWTKPG--PVLNEGSPTSIQKLVPFSKDCGPDNECVTDLVLQVNMNDIRGSKAP	814
759	Db	VDISLEN-----PGTSPALAEAYSETAKVPSIPPHKDCGEDGKICISDLVLQ--DVRIPAAQEQP	814
815	Qy	FVVRGGRKVLVSTLTLENRKENAVNTSLSIIFSNLHLASITPQRESPI---KVECC--AAP	870
815	Db	FI VSNQNKRLTFVTLKNKRRESAYNTGIVVDFSENLPFASFS-----LPVQGTVEVTCVAA	870
871	Qy	SAHARLCSVGHPVQTKAKVTFLLPEFBSGSSLLSQVFGKLTASSDSLERNGTILQENTAQ	930
871	Db	SQKSVACDVGVPALKREQQVTFTINFDFNLQNLQNAQSLSQALSESQEENKA--DNLVN	928
931	Qy	TSAYIQVEPHLLPSSSESTLHRYEVHPVGTLP-----VGPGPEPKTTLRTNNASCIVQ--	982
929	Db	LKIFLLYDABAIHLTRSTNINFIYEISSDGNVPSIVHSPFEDVGPKFIFSLKVTVTGSPVPSMA	988
983	Qy	-----NLTEPPGP-----PVHP-----BELOHTNR	1002
989	Db	TVIIHIPOYTKERNPLMYLTGVQTDKAGDICCNADINPLKIGTSSSVSPKSENFRTKE	1048
1003	Qy	LNGSNTOCQVVRCHLGOLAKGTEVSVGLLRLVHNHFFRRRAKFKSLTVVSTFELGTEEGSV	1062
1049	Db	LNCRTASCNNVTCWLKDVHMKGEYFVNVVTRIMNGTTPASSTFQTVQLTAAAEINTYNPEI	1108
1063	Qy	LQLTEASRWSSELLEVOTRPIILSLMI-----LIGSVGLGILLALLVFL	1109
1109	Db	Y-----VIEDNTVTIPLMIMKPDKEKAEVPTGVIIGSTIAGIILLALLVAIL	1154
1110	Qy	WKLGFFAHKIPBEEKLEQ	1132
1155	Db	WKLGFYKRYEKMTKKNPDEIDE	1177

## RESULT 8

```

US-08-173-497-2
; Sequence 2, Application US/08173497
; Patent No. 5437958
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van Der Vieren, Monica
; TITLE OF INVENTION: No. 5437958el Human 2 Integrin Alpha
; TITLE OF INVENTION: Subunit
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 S. Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

Db 693 KNPTLTKRRLTGLG-IHCETLKLALLPCVDEVDVSPILHLNFSLVREPIPSQNLRLPVLA 751  
Qy 773 EGSPTSQKLVPSKDCGPONECVTDVLVQNMMDIRSKAPFVVRGGRKVLVSTLEN 832  
Db 752 VGSQDLFTASLPKCNQDGLCEGD--LGVTLSFSGLOT---LTVGSSLELVIVTVWN 806  
Qy 833 RKNAYNTSLSIIFSRNLHLASLTPORESP-----IKVEC-AAPSAHARL-----CSVGHPV 883  
Db 807 AGEDSYGVTVSVLYYPAGLSHRRVSGAQKQPHQSALRLACETVPTDEGLASSRCSVNHPI 866  
Qy 884 FOTGAKTVTLLEPEFSCSSLLSQVFGKLTASSDLENGTLOENTATSAIYQEPHLF 943  
Db 867 FHESNGTFTVTFDVSYSKATLGDRM-LMRASASSENKASSKATFQLELPVKYAVVTMI 925  
Qy 944 S--SESTL-----HRYEVHPYGTLPVGPPEPKTTLRTNNASCIQNLIT 985  
Db 926 SRQESTKYFNFATSDKKKKEABHRYRVNLSQDLAISINFWVPVLLNGVA-VMDVVM 984  
Qy 986 EPPG-----PPVHPEELOHTNR---LNGSNTQCQVVRCHLQOLAKGTEVSVGLLR 1032  
Db 985 EAPQSLLPCVSEKPPQHSDFLTQISRSPMLDCSIADCLQFRCDVPSPVQEBELDTLKG 1044  
Qy 1033 LVHNEPFRRAKPSLTVSTFELCTEGSVLQLTASRWESLLE-VVOTRPILISLWIL 1091  
Db 1045 NLSFGWRETLQKVLVSVVAEITFTSVYSQLPQGFQAFMRAQMEMVLEEDVYNAIPII 1104  
Qy 1092 IGSVLGGLLALLAVFLCMLKLGFE-AHKKIPEREEKREE 1128  
Db 1105 MGSSVGALLLALITATLYLKLGFPRHYKEMLEKPED 1142

RESULT 9  
US-08-286-889-2  
Sequence 2, Application US/08286889  
Patent No. 5470953  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Mich  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE:  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: P38,659  
REFERENCE/DOCKET NUMBER: 27866/32168  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-286-889-2  
Query Match 19.0%; Score 1126; DB 1; Length 1161;  
Best Local Similarity 29.5%; Pred. No. 7.5e-99;  
Matches 348; Conservative 203; Mismatches 503; Indels 124; Gaps 41;  
Qy 23 FNLDHHPLRFPQPPPEAFEGYSVLQHVGGQRMVLGAPWDGSGDRRDGVYRCPVGAH 82  
Db 17 FNLDVEEPTIFQ-EDAGGFGQSVVQF--GGSR-LVVGAPLEVVAAQTGLYDC---AAA 69  
Qy 83 NACCAKGLHDYGLGNSSHPAVNMHGLMSLLETDDGGFMACAPLMSRACGSSVFSSGIC 142  
Db 70 TGMQCPIPL-----HIRPEAVNMSLGLTLAASNGSRLLACGPTLHRVCGENSYKSGC 123  
Qy 143 ARYDASFOPOQSLAPTAQRCF-TYMDVWIVLDCSNSI--YPMSEVOTFLRLRLVGLKPIDP 199  
Db 124 LLLGSWEIIQTVPDATPCPHQEMDIVFLIDSGSIDQNDNFQMGKGFVQAVMGQ--FEG 181  
Qy 200 EQIQVGLVQYGESPVHWSLGDFTKEEVVRAAKNLSRREGRETKTAQAIMVACTEFSQ 259  
Db 182 TDTLFAIMQVSNLLKIHFTTQFTSPSQSLVDPIVLQKGL-TFTATGILTVVTQLFHH 240  
Qy 260 SHGCRPEAARLLVVVDGESH-DGEELPAALKACEAGRVTTRYGIYVLGHVLRQRDPSSF 318  
Db 241 KNGARKSAKKILIVTDGQKYKDPLEYSQVIPAQEKAGIIRYAIGV-GHAF---QGPTA- 295  
Qy 319 LRIRITIASPPDERFFENVTDEAALTDIVDALGDRIFGLBGSNAENESSFGLBMSQIGFS 378  
Db 296 RQELNTISSAPPQDQHVKNFADNFAALGSIQKQLEKIYAVEGTOSRASSSFQHEMSQGF 355  
Qy 379 THRDKGILFGVMGAYDWGGSVLMLEGGHRLFPFRMALEDEFPPLQNLHA-----AY 430  
Db 356 TALTMGDLPLGAVGSFSW-----SGGAPLYPNMS-----PTFINMSQENVMDRDSY 402  
Qy 431 LGYSVSMLLRGRRRLFLSOAPFRIRGKVIAP-QLKKGAVRVAASLQEQGTSYFGE 489  
Db 403 LGYSTELALMKGVQNLVL-GAPRYQHTKAVITQVSQW--RKKAEVTTQIGSYFGAS 459  
Qy 490 LCPDTRDGTDTDLVLAAPMFLGPQNKETGRVYVVLV--GQOSLLTLQGTLOPEPPQD- 546  
Db 460 LCSVDVDSGSDTLILIGAPHYY--BOTRGQVSVCPPLPRQVQWQCDVLRGEQHPW 517  
Qy 547 ARFGFAMGALPDNLQDGFADVAVGAPLEDHGOALYLYHG-TQSGVRPHPAQRIAAAMP 605  
Db 518 GRFGAALTVLGDVNEKLDVGAIGAPQENRGAVLFGASGSGISPSHSQRIASSQLS 577  
Qy 606 HALSYFGRSVDGRDLDDGLDVAVGAQAAALLSRPIVHLTPSLVTPQASIVVQRD 665  
Db 578 PRLOYFQALSGGQDLTQDGLMDLAVGARGQVILLRSLPVLKVGAMRPFSEVAKAVYR 637  
Qy 666 C-----RRRQBAVCLTAALCFQVTSRTPTGRWDHQFYMRFTASLDWTAGARAAFDGS 718  
Db 638 CWEEKPSALEAGDATVCLTIQ-----KSLDQLGDIQSSVRFDLALDGLTGRAINET 692  
Qy 719 GQRLSPRLRLSVGNVTCQOLHFHVLVD-TSDYLRLPVALTVTFALDNTTKPG-----PVLN 772  
Db 693 KNPTLTKRRLTGLG-IHCETLKLALLPCVDEVDVSPILHLNFSLVREPIPSQNLRLPVLA 751  
Qy 773 EGSPTSQKLVPSKDCGPONECVTDVLVQNMMDIRSKAPFVVRGGRKVLVSTLEN 832  
Db 752 VGSQDLFTASLPKCNQDGLCEGD--LGVTLSFSGLOT---LTVGSSLELVIVTVWN 806  
Qy 833 RKNAYNTSLSIIFSRNLHLASLTPORESP-----IKVEC-AAPSAHARL-----CSVGHPV 883  
Db 807 AGEDSYGVTVSVLYYPAGLSHRRVSGAQKQPHQSALRLACETVPTDEGLASSRCSVNHPI 866  
Qy 884 FOTGAKTVTLLEPEFSCSSLLSQVFGKLTASSDLENGTLOENTATSAIYQEPHLF 943  
Db 867 FHESNGTFTVTFDVSYSKATLGDRM-LMRASASSENKASSKATFQLELPVKYAVVTMI 925  
Qy 944 S--SESTL-----HRYEVHPYGTLPVGPPEPKTTLRTNNASCIQNLIT 985  
Db 926 SRQESTKYFNFATSDKKKKEABHRYRVNLSQDLAISINFWVPVLLNGVA-VMDVVM 984

Qy 986 EPPG-----PPVHPELOHTNR---LNGSNTQCVVRCHLQGLAKGTESVSGLLR 1032  
Db 985 EAPQSPLPCVSRKPPQHSDFLTQISRSPMLDCSIADCLQPCDVPSPVQOEELDTLKG 1044  
Qy 1033 LVHNEPRRRAKFKSLTVVSTFELGTESGVLQLTASRWSESLLE-VVQTRPILISLWIL 1091  
Db 1045 NLSFGWRETLQKVLVSVVAEITFTDSVVSQLPQCAFMRQAQMEMVLEDEEVNAIPII 1104  
Qy 1092 IGSVLGGLLLALLVFLWLKLGFF-AHKKIPEEBKREE 1128  
Db 1105 MGSSVGALLLALLATATLYKLKLGFFKRYKEMLEDKPED 1142

## RESULT 10

US-08-485-618-2  
Sequence 2, Application US/08485618  
Patent No. 5728533  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103

CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago

STATE: Illinois

COUNTRY: United States

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA: US/08/485,618

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/173,497

FILING DATE: 23-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/286,889

FILING DATE: 5-AUG-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/362,652

FILING DATE: 21-DEC-1994

ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.

REGISTRATION NUMBER: 38,659

REFERENCE/DOCKET NUMBER: 27866/32797

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1161 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-485-618-2

Query Match 19.0%; Score 1126; DB 1; Length 1161;  
Best Local Similarity 29.5%; Pred. No. 7.5e-99;  
Matches 348; Conservative 203; Mismatches 503; Indels 124; Gaps 41;

Qy 23 FNLDHHPRLFPPEAEFGYSLVQHVGQGRWMLVGPWPDPGSDRRGDVYRCPVGAH 82  
Db 17 FNLDVEETIFQ-EDAGGFGQSVQVF--GGR-LVVGAPLEVAANQTRLYDC----AAA 69

Qy 83 NAPCAKHLGDYQLGNSSHPVNMHLGMSLLETDDGGGFMACPLWSRACGSSVFFSGIC 142

Db 70 TGMQOPIPL-----HTRPEAVNMSLGLTLAASTNGSRLLACGPTLHRVCGENSYSKGSC 123  
Qy 143 ARVDASFPQSQSLAPTAQRCP-TYMDVVIVLDGSNSI--YPMSEVQTFRLRLVGLKIDFP 199  
Db 124 LLLGSRWEIIQTVPDAPTECPHQEMDIVFLIDGSGSIDQDNFMQKMGFVQAVMQQ--PEG 181  
Qy 200 EQIQVGLVQVGEPSVHWSLGDFTKEEVVRAAKNLSRREGRETAKTAQAINVACTEFGSQ 259  
Db 182 TDTLFAIMOYSLNLIKIHFTFTQRTSPSQSLVDPIVOLKGL-TFTATGILTVVTQTFPHH 240  
Qy 260 SHGSRPEAARLLVVVTDGESH-DGEELPAALKACEACRGRVTRYGIAVLGHVLRDRDSSP 318  
Db 241 KNGARKSAKKILVITDGQTKDPLEYSDVIPQEKAGIIRYALGV-GHAP---QGPTA- 295  
Qy 319 LREIRTIASDPDERFFNFVTTDEAALTIDVDALDGIFLEGSHAENSSFGLEMSQIGFS 378  
Db 296 RQELNTTISSAPPQDHVFKVDNFAALGSIQKLOQEKIYAVEGTQSRASSSFQHEMSQEGFS 355  
Qy 379 THRLKDGILFGMVGAYDWGGSVLWLEGCHRLFPPRMALEDEFPFALQNH-----AY 430  
Db 356 TALTMMDGLFLGAVGFSW-----SGGAFLYPPNMS-----PTFINMSQENVDMRDSY 402  
Qy 431 LGYSVSSMLLRGGRRLFLSGAPRFRHKGVIAR-OLAKKDGAVRVAQSLQGEQIGSYFGSE 489  
Db 403 LGYSTELALWKGONLVL-GAPRYQHTGKAVITQVSRQW--RKAETVGTQIGSYFGAS 459  
Qy 490 LCPLDTRDGTVDVLLVAAPMFLGPQNKETGRVTVYLV--GQSSLLTLQGTLOPEPPD- 546  
Db 460 LCSVDVDSGSDTLILIGAPHYY--EQTRGQVSVCPPLRGQRVQWQCDVALRGEQHPW 517  
Qy 547 ARFGFANGALPDNLQDGFADVAVGAPLEDHQGALYLYHG-TQSGVRPHPAQRIAAASMP 605  
Db 518 GRFGAALTVLGDVNNEDKLIIDVAIGAPGEQENRGAVILFHGASESGISPSHQRIASSQLS 577  
Qy 606 HALSYFGRSVDGRLDLDDGLVDVAVCAQGAAILLSRPIVHLTPSLTEVTPQALSVVQRD 665  
Db 578 PRQYFGQALSGGQDLTQDGLMDLAVGARGQVLLRLSLPLVKGVAMRFSPEVAKAVYR 637  
Qy 666 C-----RRRGQEAVALCLTAALCFQVTSRTPGRWDHOFYMRFTASLDEWTAAGAAFDGS 718  
Db 638 CWEEKPSALEAGDATVCLTIQ-----KSLDLQDLGIQSSVRFDLALDGRILTSRAIFNET 692  
Qy 719 GQRLSPRLRLSVGNVTCEQLHFHVL-D-TSYLAPVALTVTFALDNTTKPG-----PVLN 772  
Db 693 KNPTLTRRKTGLG-IHCETLKLKLLPDCVEDVWSPFIILHLNFSLVREP IPSFQNLRPVLA 751  
Qy 773 EGSPTSIQKLVPSKDCGPNCEVTDVLQVNMIDIRGSRKAPFVVRGRRKVLVSTTLEN 832  
Db 752 VGSODLFTASLPFFKNCQDGLCEGD--LGVTLSFSGLT---LTVGSSLSLNVIVTVWN 806  
Qy 833 RKENAYNTSLSIIIFSRNLHLASLTPQRESP-----IKVEC-AAPSAHARL-----CSVGHPV 883  
Db 807 AGEDSYGTVVSLLYYPAGLSHRVSVGAQKHQSALRLACETVPTDEGLSRSCSVNHP 866  
Qy 884 FQTGAKVTELLPEFSCSSLLSQVFGKLTASSDLERNGLTQENTQATSAIVIQEPHLLF 943  
Db 867 FHEGNGTFTVTFDVSFKATLGDRLM-LMRASASSENKASSKATFQLELPVKVAVYTWI 925  
Qy 944 S--SESTL-----HRYEVHPYGTLPVGPPEEFTTLRTNNAACIVQNT 985  
Db 926 SRQESTKYFNFATSDKKMKEAEHRYRVNLSORDLAISINFWVPVLLNGVA-VMDVVM 984  
Qy 986 EPPG-----PPVHPELOHTNR---LNGSNTQCVVRCHLQGLAKGTESVSGLLR 1032  
Db 985 EAPQSPLPCVSRKPPQHSDFLTQISRSPMLDCSIADCLQPCDVPSPVQOEELDTLKG 1044  
Qy 1033 LVHNEPRRRAKFKSLTVVSTFELGTESGVLQLTASRWSESLLE-VVQTRPILISLWIL 1091  
Db 1045 NLSFGWRETLQKVLVSVVAEITFTDSVVSQLPQCAFMRQAQMEMVLEDEEVNAIPII 1104  
Qy 1092 IGSVLGGLLLALLVFLWLKLGFF-AHKKIPEEBKREE 1128



Db 1105 MGSSVGALLLALITATLYKLGFFKRYKEMLEDKPED 1142

RESULT 11

US-08-362-652-2

Sequence 2, Application US/08362652

Patent No. 5766850

GENERAL INFORMATION:

APPLICANT: Gallatin, W. Michael

APPLICANT: Van der Vieren, Monica

TITLE OF INVENTION: No. 576850e1 Human 2 Integrin Alpha Subunit

NUMBER OF SEQUENCES: 93

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 233 South Wacker Drive, 6300 Sear Tower

CITY: Chicago

STATE: Illinois

COUNTRY: United States

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/362,652

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/173,497

FILING DATE: 23-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/286,889

FILING DATE: 5-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.

REGISTRATION NUMBER: 38,659

REFERENCE/DOCKET NUMBER: 27866/32391

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1161 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-362-652-2

Query Match 19, 0%; Score 1126; DB 1; Length 1161;

Best Local Similarity 29.5%; Pred. No. 7.5e-99;

Matches 348; Conservative 203; Mismatches 503; Indels 124; Gaps 41;

Qy 23 FNLDHEHPRIFFPGPEAEFGYSVLQHVGGQRMVLGAPMDPGSDRRGDVYRCPVCGAH 82

Db 17 FNLDVEPTFIQ-EDAGGFGSVQF--GGSR-LVVGAPLEVVAANTQGRLYDC---AAA 69

Qy 83 NAPCAKHLGDIYQGNSSHPVAVNHGLMSLLETDDGGFWACAPLMSRACSSVFFSGIC 142

Db 70 TGMCOPIPL-----HIRPEAVNMSLGLTLAASTNGSRLLACGPTLHRVCGENSYSKGSC 123

Qy 143 ARVDASFPQGSGLAPTAQRCP-TYMDVVIVLDGNSI--YPMSEVQVFLRLVGLKFLDP 199

Db 124 LLLGSRWEIITVPDAPTECPHQEMDVFLLIDGSGSIDQDNFNQMGFVQVMQ--FEG 181

Qy 200 EQIQVGLVQYGESPVHWSLSDFRKKEEVVRAAKNLSRREGRETAKTAQAINVACTEGFSQ 259

Db 182 TDTLPAWMQYNLKKIHTFTQFRTSPSQSLVDPIVLQKGL-TFTATGILTVVTQLFHH 240

Qy 260 SHGRPEARLLVVVTDGESH-DEEELPAALKACEAGRVTRYGTAVLGHYLRORDPSF 318

Db 241 KNGARKSAKKLIVITDQKYKDPLEYSVDVIPQAEKAGIIRYAGV-GHAF---QGPTA- 295

Qy 319 LREIRTIASDPDERFFNFVNTDEAALTDIVDALGDRIFGLEGSHAENESSFGLMSQIGFS 378

Db 296 RQELNTISSAPPQDHFVKVDNFAALGSIQKQLEKIYAVEGTQSRASSSFQHEMSQEGFS 355

Qy 379 THRLKDGILFGMAYDWDGSGVLMLEGGHRLFPFPRMALEDEFPALQNH-----AY 430

Db 356 TALTMGDLFLGAVGFSW-----SGGAPLYPNMS-----PTFINMSQENVMDRDSY 402

Qy 431 LGYSVSSMLRGRRRLFLSGAPRHRGKVIAF-QLKKGAVRVAQSLQEQIGSYFGSE 489

Db 403 LGYSTELALWKGVQNLVL-GAPRIQHTGKAVITQVSROW--RKKAEVGTQIGSIFGAS 459

Qy 490 LCPLDTRDGTDTVLVLAAPMFLGPNQKTKETGRVYVYL--GQOSLLTLQGTLPPEPQD- 546

Db 460 LCSVDVDSGSDTLILIGAPHY--BQTRGGQSVSCPLPRGQVQVQCDVLRGEQHPW 517

Qy 547 ARFGFANGALPDNLQDGFADVAVGAPLEDHGHQALYLYHG-TOSGVRPHPAQRIAAAMP 605

Db 518 GREGAALTVLGDVNEKLDLIDVAIGAPGEQENRGAVYLFHGASRSGISPSHSQRIASSQLS 577

Qy 606 HALSYFGSRVDGRDLDDGDLVDVAVGAGAAALLSSRPVHLTPSLVTPQAISVVQRD 665

Db 578 PRLQYFGQALUSGGDLTQDGLMDLAVGARGQVLLSLPLVKVGVAMRFSFVEKAVYR 637

Qy 666 C-----RRRQGEAVCLTAALCFQVTSRTTPGSRWDHQFYMRFTASLDEWTAGARAAFDGS 718

Db 638 CWEEKPSALEAGDATVCLTIQ-----KSSLDQLDGIQSSVRFDLALDPGLRITSRAINFET 692

Qy 719 GQRLSPRRLSVGNVTCBOLHFFHVD-TSDYLRPVALVTVPALDNTKFG-----PVLN 772

Db 693 KNPTRLRRKTLGLG-IHCETLKLLPDCVEDVVSPIILHLNFSLVREPIPSPNQRLRPVLA 751

Qy 773 EGSPTSIOKLVPFSKDCGPDNECVTLVLQVNMDIRSKKAPFVVRGGRKVLVSTTLEN 832

Db 752 VGSQDLFTASLPFEKNCGQGLCEGD--LGVTLISFSGLQT---LTVGSSLELNVIVTVN 806

Qy 833 RKENAYNTSLIIFSRNLHLASLTPORES-----IKVEC-AAPSAHARL---CSVGHPV 883

Db 807 AGEDSYGTVVSLLYYPAGLSHRVSGAQKQPHQSALRALACETVTEDEGLSSRCSVNHP 866

Qy 884 FQTGAKVTLLRPFECSSLLSQVFGKLTASSDSLRNGTLQNTAGTSYIYQEPHLFF 943

Db 867 FHEGNGTFTVTFDVSFKATLGDRM-LMRASSENKASSKATFQLELPVYAVYVIMI 925

Qy 944 S--SESTL-----HRYEVHPYCTLVPGCPPEKTTTLRTNASCIVQNLT 985

Db 926 SROEESTKYPNFATSEKMKAEHRYRVNLSQDLAISINFVWPVLLNGVA-VMDVVM 984

Qy 986 EPPG-----PPVHPELQHTNR---LNGSNTQCQVVRCHLQLAGTEVSVGLLR 1032

Db 985 EAPSQSLPCVSEKPPQHSDFLTQISRSPMLDCSIADCLQFRCDVPSFSVQEELDTLKG 1044

Qy 1033 LVHNEFPRAKFKSLTVVSTFELGTBEGSVLQTEASRWSESLLE-VVQTRPLISLWIL 1091

Db 1045 NLSFGMWRETQLQKVLVVAEITFDTSVYSQLPQGEAFMRAQMEMVLEEDVYNAIPII 1104

Qy 1092 IGSVLGGLLLALLVFCIMKLGFF-AHKKIPSEKREE 1128

Db 1105 MGSSVGALLLALITATLYKLGFFKRYKEMLEDKPED 1142

RESULT 12

US-08-605-672-2

Sequence 2, Application US/08605672

Patent No. 5817515

GENERAL INFORMATION:

APPLICANT: Gallatin, W. Michael

APPLICANT: Van der Vieren, Monica

TITLE OF INVENTION: No. 581751e1 Human 2 Integrin Alpha Subunit

NUMBER OF SEQUENCES: 103

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:52 ; Search time 48.3117 Seconds  
(without alignments)  
1749.117 Million cell updates/sec

Title: US-09-647-544-4  
Perfect score: 5932  
Sequence: 1 MELPFVTHLFLPLVFLTGLC.....GFFAHKKIPBEKREKLEQ 1132

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES \*

Result No.	Score	Query Match	Length	ID	Description
1	2345	39.5	1217	4	US-09-949-016-7892
2	1884	31.8	1180	4	US-09-000-004A-2
3	1775.5	29.9	1183	4	US-09-532-310B-5
4	1733.5	29.2	1181	4	US-09-000-004A-4
5	1733.5	29.2	1181	4	US-09-949-016-6189
6	1733.5	29.2	1195	4	US-09-949-016-10747
7	1704.5	28.7	1183	4	US-09-532-310B-6
8	1126	19.0	1161	1	US-08-173-497-2
9	1126	19.0	1161	1	US-08-286-889-2
10	1126	19.0	1161	1	US-08-485-618-2
11	1126	19.0	1161	1	US-08-362-652-2
12	1126	19.0	1161	2	US-08-605-672-2
13	1126	19.0	1161	2	US-08-482-293A-2
14	1126	19.0	1161	2	US-08-943-363-2
15	1126	19.0	1161	3	US-09-193-043-2
16	1126	19.0	1161	4	US-09-688-307A-2
17	1126	19.0	1161	4	US-09-350-259-2
18	1124	18.9	1161	3	US-09-193-043-55
19	1124	18.9	1161	4	US-09-688-307A-55
20	1124	18.9	1161	4	US-09-350-259-55
21	1120.5	18.9	1161	1	US-08-485-618-99
22	1120.5	18.9	1161	2	US-08-605-672-99
23	1120.5	18.9	1161	2	US-08-482-293A-99
24	1120.5	18.9	1161	2	US-08-943-363-99
25	1120.5	18.9	1161	3	US-09-193-043-99
26	1120.5	18.9	1161	4	US-09-688-307A-99
27	1120.5	18.9	1161	4	US-09-350-259-99

RESULT 1  
US-09-949-016-7892  
; Sequence 7892, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7892  
; LENGTH: 1217  
; TYPE: PRT  
; ORGANISM: Human  
; US-09-949-016-7892

## ALIGNMENTS

28	1118	18.8	1161	1	US-08-485-618-55	Sequence 55, Appl
29	1118	18.8	1161	1	US-08-362-652-55	Sequence 55, Appl
30	1118	18.8	1161	2	US-08-605-672-55	Sequence 55, Appl
31	1118	18.8	1161	2	US-08-482-293A-55	Sequence 55, Appl
32	1118	18.8	1161	2	US-08-943-363-55	Sequence 55, Appl
33	1115.5	18.8	1170	2	US-08-789-078-2	Sequence 2, Appl
34	1115.5	18.8	1170	2	US-08-752-633-2	Sequence 2, Appl
35	1115.5	18.8	1170	5	PCT-US95-04886-2	Sequence 2, Appl
36	1114.5	18.8	1161	1	US-08-485-618-53	Sequence 53, Appl
37	1114.5	18.8	1161	1	US-08-362-652-53	Sequence 53, Appl
38	1114.5	18.8	1161	2	US-08-605-672-53	Sequence 53, Appl
39	1114.5	18.8	1161	2	US-08-482-293A-53	Sequence 53, Appl
40	1114.5	18.8	1161	2	US-08-943-363-53	Sequence 53, Appl
41	1114.5	18.8	1161	3	US-09-193-043-53	Sequence 53, Appl
42	1114.5	18.8	1161	4	US-09-688-307A-53	Sequence 53, Appl
43	1114.5	18.8	1161	4	US-09-350-259-53	Sequence 53, Appl
44	1108.5	18.7	1155	1	US-08-286-889-46	Sequence 46, Appl
45	1108.5	18.7	1155	1	US-08-485-618-46	Sequence 46, Appl

Query Match 39.5%; Score 2345; DB 4; Length 1217;  
Best Local Similarity 42.2%; Pred. No. 1.3e-216;  
Matches 505; Conservative 204; Mismatches 407; Indels 80; Gaps 16;

Qy	1	MELPFVTHLFLPLVFLTGLCSPFNLDHHPPLFPPEAEFGSVLQHVGGQRMVLVGA	60
Db	30	MDLPRGLVVAWALSILWPGFTDFNMDTKPRVIFGSRTAFTGYTVOQHDISGNKWLVGA	89
Qy	61	PWDGPGDRDRYVRCVPVGGAHNAPCAKHGLDGYQLGNSSHPANMHLGMSLLETDDGG	120
Db	90	PLETNGYQKTGDVYKCPV---IHGNCCTKLNGLRVTLNSVSRKDNMRLGLSLATNPKN	146
Qy	121	FMACAPLWSRACGSSVFSSGICARVDASFPQGSIAPTAOCPTVMVIVLDGNSIYP	180
Db	147	FLACSPILWSHCEGSSYTTGMCNRVNSFRFSKTVAPALQRCQTYMDIVIVLDGNSIYP	206
Qy	181	WSEVQTFRLRLVGLKFLIDPEQIQVGLVOYQSPVHWSLGDGFRTEKEEVVRAAKNLSR	240
Db	207	WVEVQHFLINILKFFYICPGQIQGVVQYQGVGVHFEHLNDYRSKDVVVAASHLEQRCG	266
Qy	241	RETKTAQIMVACTGFSQSHGSRPEARLIVVTVTDGSHDGEELPAALKACEAGRVTY	300
Db	267	TETRTAFGIEFARSEAFQK--GGRKGAKKVMIVITDGESHSDSPDLEKVIQSSERNV	324
Qy	301	GIAVLGHYLRQRDPSSFLREIRTIASDPDRFPFNVTVDEAALTDIVDALGDRIFGLSGS	360

```
Db 325 AVAVGYNRRGINPETFLEIKYIASDPDDKHFFNVTDEAALKDIVDALGDRIFSLEGT 384
Qy 361 HAENESSFGLSEMSOIGFSTHRLKDGILFGMVAGVDMGSGVLMLEGGHRLFPFPMALDEF 420
Db 385 N-KNETSFGLEMSOTGFSSHVVEDGVLVAGVAGVDMGAVLAKETSAGKVIPLRESYLKEF 443
Qy 421 PPALQNHAAVLGYSVSSMLRGGRRRLFLSGAPRFRHRGKVIAFOLKKDGAVRVAQSLQGE 480
Db 444 PEELKNGHVLGYITVTSVSSRQGRVYVAGAPRHNHTGKVLFTMHNRSITLHQMRGQ 503
Qy 481 QIGSYFGESELCPLDTRDGTDLVLLVAAPFLGPQNKETGRVYVYLVGQOSLLTLQCTLQ 540
Db 504 QIGSYFGESEITSDIDGTDVLLVAGAPMYFN-EGRERGKVVYVEL-RQNRFYNGTLK 561
Qy 541 PEPP-QDAREGFAMGALPDNLNOGFADVAVGAPLEDHOGALYLYHGTQSGVRPHRARI 599
Db 562 DSHSYQNRFGSSIASVRDLNQSNDVVGAPLEDNHAGAIYIFHGFSGSILKTPKORI 621
Qy 600 AAASMPHALSYFGRSVDRGLDLDDGLVDVAVGAQGAAILSSRPVHLTPSLEVTPOAI 659
Db 622 TASELATGLQYFGCSIHGQLDDEDGLDLAVGALGNVILMSRPVQINASHLFEFSKI 681
Qy 660 SVVORDCRRGQEAVALCTAALCFQVTSRTPGRMDHQYMFRTASLDWTAGARAAFGSG 719
Db 682 NIFHRDCKRGRDATCLAAFLCFPTPIFLAPHFQTTVGIRYNATMDERRVTPRAHLDEGG 741
Qy 720 QRLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALTVTTFALDNTTKPGPVLINEGSPSI 779
Db 742 DRFTNRVALLSSGQELCERINFHVLDTADYVKVTFVSVEYSLEDPDH-GPMLDDGWPTTL 800
Qy 780 QKLVFPKDCQPDNECVTDLVQVNMDI-----RGSRK-----APFVV 817
Db 801 RVSVFPWNGCNEDEHCVDPDLVDARSDLPTAMEYCORVLRKPAQDCSAYTILSPDTTVFII 860
Qy 818 RGGRRKVLVSTTLNRKENAYNTSLTIFGRNLHLASLTTPQRESPIKVECAAPS--AHAR 875
Db 861 ESTQRVAVEATLENRGENAYSTVLNISQSANLQFASLIQKEDSDGSEICVNEERRLOKQ 920
Qy 876 LCSVGHVFTQGAFTLLPEPSCSLLSOVFOKLTASSDSLSRNGTLQENTTAQTSAYI 935
Db 921 VCNVSYFFRAKAVAFRLDPFEFSKIFLHLELELAAGSDSNRDKEDNVAPLRFHL 980
Qy 936 QYEPHLLFSSBESTLHRYEVHPYGLP--VGPGPEFKTLRTN----- 975
Db 981 KYEADVLPTRSSLSHVEKLNSSLERYDGTGPPFCIFRIONLGLFPIHGMWMMKIPI 1040
Qy 976 -----NASC-IVQNLTEPPGPPVHPPELOHTNRLNGSNTCCQVVR 1015
Db 1041 ATRSGNRLKLRLDPLTDEANTSCNIGNSTEYRTPVE-EDLRRAPQLNHSNSDWSINC 1099
Qy 1016 HLGOLAKGTVSGLLRVHNEPFRRAKFKSLTVVSTFELGTBEGSVLQLTEASRWSESL 1075
Db 1100 NI-RLVPNQEFNHLGNLMRLSKALKYKSMKIMVNAALQRPSPHFFIFREEDPSRQIV 1158
Qy 1076 LEVVOTRPLISLWILGSLVGLLALLVFLCWLKGLGFFAHKKIPEEKREKLE 1131
Db 1159 FEISKQEDWQPIWIIIVGSTLGGILLALLVALLKWLKGFPRSA-----RRREPGLD 1210

RESULT 2
US-09-000-004A-2
; Sequence 2, Application US/090000004A
; Patent No. 6780603
; GENERAL INFORMATION:
; APPLICANT: Tsalibary, Photini-Effie
; APPLICANT: Charonis, Aristidis S.
; APPLICANT: Setty, Suman
; APPLICANT: Mauer, Michael
; TITLE OF INVENTION: ANALYSIS OF ALPHA INTEGRINS FOR THE DIAGNOSIS OF DIABETIC NEPHRO
; FILE REFERENCE: 600.314USWO
; CURRENT APPLICATION NUMBER: US/09/000,004A
; CURRENT FILING DATE: 2001-05-02
```

```
; PRIOR APPLICATION NUMBER: US 60/001,387
; PRIOR FILING DATE: 1995-07-21
; PRIOR APPLICATION NUMBER: US 60/001,861
; PRIOR FILING DATE: 1995-08-03
; PRIOR APPLICATION NUMBER: US 60/016,700
; PRIOR FILING DATE: 1996-05-02
; PRIOR APPLICATION NUMBER: PCT/US96/12067
; PRIOR FILING DATE: 1996-07-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 1180
; TYPE: PRT
; ORGANISM: Rattus
US-09-000-004A-2

Query Match 31.8%; Score 1884; DB 4; Length 1180;
Best Local Similarity 36.2%; Pred. No. 4,4e-172;
Matches 432; Conservative 219; Mismatches 432; Indels 112; Gaps 25;

Qy 13 LVFLTGLCSPFNLDHHPRLFCPPPEAEFGYSLQHVGGQRMLVGCAPWDGPGSGRRGD 72
Db 19 LTVLGFCSFNVVDVXNSMSFGPVEDMFGYTVQQYENEBGKWLVLGSLVGQPKARTGD 78
Qy 73 VYRCPVGGAHNAPCAKGHLDGYQLGNSSHPAV-----NMHLGMSLLLETDCGGFMACAPL 127
Db 79 VYKCPVGRERAMPCKLDLP---VNTSIPNVTEIKENMTFG-STLVNPNNGGFLACGPL 133
Qy 128 WSRACSSVFPSSGICARVDASFOQCSLAPTAQRCPTYMDVIVLDGNSIYIPWSEVQTF 187
Db 134 YAYRCHLHYTTGICSDVSPTFQVNSFAP-VQECSTQLDIVILVLDGNSIYIPWSEVIAF 192
Qy 188 LRRLVGLKFLDPQIOVGLVQYGESPVHMSLGDFTKEBVRRAAKNLSRREGRETAKQ 247
Db 193 LNDLLKRMIDIGPKQTQVIGVQYGENTHFNLNKKYSTEEVLVAANKIGQGLQNTAL 252
Qy 248 AIMVACTEGFSQSGHGRPEAARLLVVVTDGESHDBELPAALKACAGRVTRYGIAVLGH 307
Db 253 GIDTARKEAFTEARGARRGVKVMVITVDTGESHDRYLRKQVIOCEDENIQRFSAIILGH 312
Qy 308 YLRQRDPSSFLREITIASDPPERFPFNVTDEAALTDIVDALGDRIFLGEGSHAENESS 367
Db 313 YNRGNLSLTKFVEBEIKSIASEPTEKHFNFVSDSLALVTIVKALGERIFALEATADQSAAS 372
Qy 368 FGLEMSOIGFSTHRLKDGILFGMVAGVDMGSGVLMLEGGHRLFPFPMALDEFPPALQNH 427
Db 373 FEMESOTGFSALYSQDWMVLMGAVGAYDMNGTVVMQKANKQNVIPHNTTFQTE--PAKONE 430
Qy 428 --AAYLGYSVSSMLRGGRRRLFLSGAPRFRHRGKVIAFOLKKDGAVRVAQSLQGEQIGSY 485
Db 431 PLASYLGYTVNSATIPGD-VLYIAGQPRYNHTGQVVIYKM-EDGNINILQTLGGEQIGSY 488
Qy 486 FGSELCPLDTRDGTDLVLLVAAPFLGPQNKETGRVYVYLVGQOSLLTLQCTLOP---- 541
Db 489 FGSLVTLTIDIDKOSYDILLVAGAPMYMGTEKBEQKGVYVYAV-NQTRFBYQMSLEPIROT 547
Qy 542 -----EPQDAREGFAMGALPDNLNOGFADVAVGAPLEDHOGALYLYHG 586
Db 548 CCSSLLKONSCTKENKNEPCGARFGTAIAVKDLNVDFNDVVICAPLEDHAGAVIYHG 607
Qy 587 TQSGVRPHRARIAAAAMPHALSIFGRSVDGRDLDDGLVDVAVGAQGAAILSSRPV 646
Db 608 SGKTIREAYAIRIPSGGDGKTLAKFFGQS IHGENDLNGDGLTDTVTIGGLGAALFWARDVA 667
Qy 647 HLTPSLEVTPOAISVVORDCRRGQEAVALCTAALCFQVTSRTPGRMDHQYMFRTASLDE 706
Db 668 WKVTWNFEENKVNIOKKNCRVEGKETVCVTCINATMCFHVKLKSKEDSYEADLQYRVTLDS 727
Qy 707 WTAGARAAFGSGGRSLSPRLRLSVGNVTCEQLHFHVLDTSDYLRPVALTVTTFALDNTTK 766
Db 728 LRQISRFSGTGQERKIQR--NITVRESECIHSIFMLDKHDFQDSVRVTLDF---NLTD 782
Qy 767 P--GPVLNEGSPTSIQKLVFPKDCQPDNECVTDLVQVNMDIRSGRKAPFVVRGRRKV 824
```

Db 783 PENCPLDLPNSVHEIIPAKCGNKERCISDLTLNVST-----TEKSLILVKSQHDKF 838  
Qy 825 LVSTTLNKRKENAYNTSLSIIFSRNLHLATLP-QRESPIKVECAAPSAHARLCSVGHVP 883  
Db 839 NVSLTVNKGDSAYNTRTVQHSPLNIFSGIEEIQDS-----CESNQNTCRVGYPP 891  
Qy 884 FOTGAKVTFLLEPFSCSSLLSQVFGKLTASSDSLRNGTLOENTAOISAYIQEPHLLF 943  
Db 892 LRAGETVTFKIFQFNTSHLSENAIHLSATSDSEEPLESUNDNEVNIPIPKYEVGLQF 951  
Qy 944 SSESTLHRYEHPYGTLP-----VG-----PGPEFKTL----- 972  
Db 952 YSSASEHHISVAANETIPEFINSTEDIGNEINVPYTIKRGHPMPPELQLSIFPNLTAD 1011  
Qy 973 -----RTNNASCIQNLTPEPG-----PPVHPEELOHTNRLNGSNTQCVVRC 1015  
Db 1012 GYPVLYPIGWSSDNVNCPRSLDPFGINSCKMTSKSVLKRGTIQDCSSTCGVATI 1071  
Qy 1016 HLGOLAKG-TRVSGLLRLVHNEFRRAKFKSLTVVSTFELGTGEGSVLQLTASRWSES 1074  
Db 1072 TCSLLPSDLSQVNSL--LWKPTFIRAHFSSNLTLRGELKSENS--LTLSSNRKREL 1128  
Qy 1075 LLEVQOT-RPTLISILWILGVLGGLLALLVFLCKLWKGFPFAHKIPEEEKREE 1128  
Db 1129 AIQSKDGLFORVPLWILLSAPAGLLMLLMLLALWKIGF---KRPLKKMEK 1180

## RESULT 3

US-09-532-310B-5  
Sequence 5, Application US/09532310B  
Patent No. 6596276

## GENERAL INFORMATION:

APPLICANT: Senger, Donald R  
Detmar, Michael  
Claffey, Kevin P

TITLE OF INVENTION: Method for inhibiting tumor  
angiogenesis in a living subject

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: David Prashker, Esq.

STREET: P.O. Box 5387

CITY: Magnolia

STATE: Massachusetts

COUNTRY: USA

ZIP: 01930

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.40 Mb storage

COMPUTER: Dell PC

OPERATING SYSTEM: MS DOS

SOFTWARE: Microsoft Word version 97

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/532,310B

FILING DATE: 22-Mar-2000

CLASSIFICATION: Unknown

ATTORNEY/AGENT INFORMATION:

NAME: David Prashker, Esq.

REGISTRATION NUMBER: 29,693

REFERENCE/DOCKET NUMBER: BIS-036

TELECOMMUNICATION INFORMATION:

TELEPHONE: (978) 525-3794

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1183 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-532-310B-5

Query Match

Best Local Similarity 29.9%; Score 1775.5; DB 4; Length 1183;

34.8%; Pred. No. 1.4e-161;

Matches 416; Conservative 219; Mismatches 449; Indels 111; Gaps 24;  
Qy 13 LVFLTGLCSFNLDEHHRLPFGPPEABFQSVYLQHVGGGRRMLVAGAPDGPDSGRGD 72  
Db 21 LTVILGFCVFNVDVKMSMTFSGPVEDMFGVTVQOYENEEGKWVLIGSPLVGQPKNRGT 80  
Qy 73 VYRCPVGAHNAACAHLGDYQLGNSSHPAV-----NMHLGMSLLTDDGGGFMACAPL 127  
Db 81 VYKCPVGRGSLPCVKLDLP-----VNTSIPNVTVEKENMTFG-STLVNPNGGFLACGPL 135  
Qy 128 WSRACGSVSSSGTCARVDASFOQGS LAPTAQRCPTYMDVIVLDGNSNIYPSEVQTF 187  
Db 136 YAYRCGHLHTTGICSDVSPTFQVNSIAP-VQESCXTKLDIVIVLDGNSNIYPMSDV-TA 193  
Qy 188 LRLVLGKLFDPQIQVLQVYGESPVHWSLGDFTKKEEVRRAAKNLRSREGRETAKQ 247  
Db 194 LNDLLKEMDLPKXKTVGIVXGENVTHEFNLNKYSSTEELVAAKKIVXRGGRXTMTAL 253  
Qy 248 AIMVACTEGFSQSHGGRPEAARLLVVVTDGESHDBELPAALKACAGRVTRYGVIAVLGH 307  
Db 254 GTDTARKEAFTEARGARGVKVMVIVTDEHDXNHLKKVIGDCEDENIXRFSIALUGS 313  
Qy 308 YLRRQDPSSFLREIRTIASDPDERFFFNVTDEAALTDIVDALGDRIFGLEGSHAENESS 367  
Db 314 YNRGNLSTEFVEIKSIASEPTKSFNVSDELALVTIVKTGERIFALEATADXSAAS 373  
Qy 368 FGLEMSQIGFSTHRLKDGILFGMVGAYDWGGSVLMLEGGHRLFPPRMALEDEFPPALQNH 427  
Db 374 FEMMSQTGFSAHYSQDMVWLGA VGADWNGTVVYNQASQIIIPRNTTFNVESTKQNEPL 433  
Qy 428 AAVLGYSVSMLLRGGRRLFLSGAPRFRHGRKVIATFOLKXKDGAVRAQSGOGEIGSYFG 487  
Db 434 ASYLGYTVNSATASSGDVLYIAGQPRYNHTGQVYIYRM-BEGNIKIQTLSGXIGSYFG 492  
Qy 488 SELCPLDTRDGTDLVLLVAAPMPLGPONKETGVYVYLVGQQSLLTLQGTLOP-EP--- 543  
Db 493 SILTTTIDDKDSNTNILLVGAPMYMGTEKEGGKVYVYAL-NQRFYQMSLAPMEPIKQ 551  
Qy 544 -----PQDARPFAMGALPDUNQDGFADVAVGAPLEDHGOHALYLYH 585  
Db 552 TCCSSRQHNSCTTENKNEPCGARGTAIAAVKOLNLDGDNFIVIGAPLEDHGGAVYIYH 611  
Qy 586 GTQSGVRPHPAQRITAAASMPHALSYFGRSVDGRLLDGLDLDVAVAGQAAILLSRPI 645  
Db 612 GSGKTIKKEYAXRIPSGDGKTLKFGQSHGEMDLNGDGLTDVTICGLGAALFWSRDV 671  
Qy 646 VHLTPSLEVTPOAISVVQDRCRRRQEAVALCTAALCFQVTSRTPCRDWHQFMRFTASLD 705  
Db 672 AVVKVTWNPENKVNIOKKCHMEGKETVCINATVCFEVLKSKEDITYEADLQYRVTL 731  
Qy 706 EWTAGARAAFDGSGQRLSPRRLLSVGNVTCEQLHFLHVLDTSDYLRPVALVTTFALDNTT 765  
Db 732 SLRQISRSFSGTQERKVQR-NITVRKSECTKSHFMYLMDKHDPQDSVRITLDF---NLT 786  
Qy 766 KP--GPVLNEGSPISIOKLVPFSKDCGPDNECVTDVLQVNMDIRGSRKAPFVVRGGRK 823  
Db 787 DPENGPVLDLPSLNSVHEIYTPFAKDCGKCKCISDLSLHV-----ATTEKDLIVRSQNDK 842  
Qy 824 VLVSTTLLENKENAYNTSLSIIPSRNLHLASLTPORESPIKVECAAPSAHARLCSVGHVP 883  
Db 843 FNVSLTVNKTGDSAYNTRTVHYSPLNVFSGI-----EAIQKDC--FSNNHITCKVGYPP 896  
Qy 884 FOTGAKVTFLLEPFSCSSLLSQVFGKLTASSDSLRNGTLOENTAOISAYIQEPHLLF 943  
Db 897 LRRGEMVTFKLPQFNTSYLMENVTIYLSATSDSEEPETLSDDNVNISIIPVKYEVGLQF 956  
Qy 944 SSESTLHRYEHPYGTLP-----VG-----PGPEFKTL----- 972  
Db 957 YSSASEYHISIAANETVPEVINSTEDIGNEINIFYLIRKSGSFPMPPELKLISFPNMTSN 1016  
Qy 973 -----RTNNASCIQNLTPE-----PPVHPEELOHTNRLNGSNTQCVVRC 1015  
Db 1017 GYPVLYPTGLSSSENANCRPHIFEDPFSINSKGKQNTSTDLHLKRGTLDCNCKFAITC 1076

```

QY 1016 HLCQLAKGTVEV-VGLLRLVHNEFFRRAPKFSUTVVSTELGTBEGSVLQLTASRWSS 1074
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1077 NL-----TSSDISXVNVSLILWKPTFKSYFSSNLNLTIRGELRSNAS-LVLSSENEKREL 1131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1075 LLEVQVT-RPILSLMLIGSVLGGILLALLVFCILWKLGFFPAHKKIPKEEKREE 1128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1132 AIQISKDGLGRVPLMWILLISAFAGLLMLLLILALWKIGFF---KRLPKKRMK 1183

RESULT 4
US-09-000-004A-4
; Sequence 4, Application US/09000004A
; Patent No. 6780603
; GENERAL INFORMATION:
; APPLICANT: Tsilibary, Photini-Effie
; APPLICANT: Charonis, Aristidis S.
; APPLICANT: Setty, Suman
; APPLICANT: Mauer, Michael
; TITLE OF INVENTION: ANALYSIS OF ALPHA INTEGRINS FOR THE DIAGNOSIS OF DIABETIC NEPHRO
; FILE REFERENCE: 600.314USWO
; CURRENT APPLICATION NUMBER: US/09/000,004A
; CURRENT FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/001,387
; PRIOR FILING DATE: 1995-07-21
; PRIOR APPLICATION NUMBER: US 60/001,861
; PRIOR FILING DATE: 1995-08-03
; PRIOR APPLICATION NUMBER: US 60/016,700
; PRIOR FILING DATE: 1996-05-02
; PRIOR APPLICATION NUMBER: PCT/US96/12067
; PRIOR FILING DATE: 1996-07-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 1181
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-000-004A-4

Query Match          29.2%   Score 1733.5;   DB 4;   Length 1181;
Best Local Similarity 33.7%   Pred. No. 1.6e-157;
Matches 410;   Conservative 217;   Mismatches 431;   Indels 159;   Gaps 32;

QY 11 LPLVFTLGL-----CSPENLDEHPRLPPGPEAEFGYSVLQHVGGGQRMVLGAPWD 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11 LPLLLVLALSGIILNCLLAYNVLPEAKISFGSSBQPGYAVQQFINPKGNWLLVSPWS 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 64 GPSGDRRGDYVRCPVGGAHNAPCAKHL- GDYQLGNSSHPAVNMHLGMSLLETGDDGGFM 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 GPENRMGDVYKCPV-DLSTATCEKLNLTSTGIPNVTEKTNKNSLGLILTRNMGTTGFL 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 ACAPLWRACGSSVFPSSGICARVDASPOQGS LAPTAQRCPTYMDVVVLVDGNSIYPWS 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 TCGPLMAQQCGNQYTTTGVGCSDISPDQLSASFSPATQPCFSLIDVVVVVDCESNIYPWD 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 EVOTFLRLVKGFLFIDPEQIQVGLVOYGSBPVHWSLGDPTKEEVVRAAKNLSRREGRE 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 190 AVKNFLEKFGVGLDIGNTKTQVGLIQYANNPRVVFNLTYYTKKEIMIVATSTQSYGGDL 249
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 243 TKTAQAIMVACTBGFSGSHGRPEAPARLLVVVTDGESHGDEELPAALKACEAGRVTYGI 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 250 TNTFGAIQYARKVAYSAAAGRRSATKVMVVVTDGESHGDSMLKAVIDQCNDHNDILRFGI 309
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 303 AVLGHYLRQRDDSSFLREITRTASDPDRFPFNVTDAAALTDIVDALGDRIFGLEGSHA 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 310 AVLGYLNRNALDTKNLKEIKATIASIPTEFYFNVSDEAAALLEKAGTLGQIFSIEGT-V 368
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 363 ENSSFCLEMSIQIGFST--HRLKXGILFGVMGVADWGGSVLW-LEGGHRLFPFPPMALEDE 419
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 369 QGGDNFQWMSQVGFSDYDSQNDILMLGAVGAFGWSGTIVQKTSHGHLFF-----KQA 423
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 420 FPPALQ--NHAAYLGYSVSSMLJFGGRRLLFSGAPRFRHRGKVIAFOLKXKDGAVRVAQSL 477
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

424	Db	424	FOILQDRNHSSVYGSVAA--ISTGSETHFVAGAPPANTYTGQIVLYSVNBNGNITVIQAH	482
478	Qy	478	QGEIGSYSGELCPDLTDRTDGTDTVDLLVAAAPMFLGPQNKQTCRVVYV-----LVGQQSL	532
483	Db	483	RGQIGSYSGVLCVVDVKDTITDVLVGAAPMYSMDLKKKEGRVYVFTIKKIGLGHQHF	542
533	Qy	533	LTLQGTVLQPEPPQDAPFGFAMGALPOLNOGDFADVAAGAPLEDGHOGALYLYHGTQSGVR	592
543	Db	543	--LEG---PEGIENTRFGAIAALSINNDGFDNDVTGSPLENQNSGAVYVYNGHQTIR	597
593	Qy	593	PHPAQRIAAA--SMPHALSVFGESVDGRLLDGGDLVDVAVGAQGAAILLSRPVHLTP	650
598	Db	598	TKYSQKILSGDGAFRSHLQYFGSLDYGDLNGDSITDVSIGAGVQVQWMSQSIADVAI	657
651	Qy	651	SLEVTPOAITSVVORDCCRRGQEAIVCLTAALCFQVTSRTTGRMDHQFYMFRTASLDEWTAG	710
658	Db	658	EASFTEPKITLVNKNQA-----IILKLCFSAKER-PTKQNNQVAIVYNTILDA----	704
711	Qy	711	ARAAFGSGQRISPRRL-----RLSGNV-----TCQLHPHVLDTSDYLRPVALTVT	758
705	Db	705	-----DGFSRVTSRGLFKENNERCLQKNMVVNAQSCPEHIITYIQPSDVVNSLDLRVD	759
759	Qy	759	FALDNTTKG--PVLNEGPTSLQKLVPSKDCGPDNECVTDVLQVQNDMDIRSGRAFPV	816
760	Db	760	ISLEN---PQTSALBAYSETAKVFSIPPHKOCGEDGLCISDLVLDVR-QIPAAQSQPFI	815
817	Qy	817	VRGGRKVLVSTTLENKKNAYNTSLSIIFSRNLHLASLTPQRESPI---KVBC-AAPSA	872
816	Db	816	VSNQKRLTFSTVLKKNRESAYNTGIVDFSENLPFASP-----LPVDGTEVTCQVAASQ	871
873	Qy	873	HARLCSVHPVFTQGAKVFTLLLEPFSCSLLSQVFGKLTAGSDSLERNGTQLQNTAQT	932
872	Db	872	KSVACDVGPALKRQEQVFTINFDNLQNLQNASLSFQALSESQENKA--DNLVNLK	929
933	Qy	933	AYIQEPHLLFSESSTLHYEVHPYGTLP-----VGPGEFXTTIRTNWASCIQV----	982
930	Db	930	IPLLYDAEHLTRSTINIFYEISDGNVPFSIVHSFEDVGPKFIISLKVTTGTVFVSMATV	989
983	Qy	983	-----NLTPPGP-----PVHP-----EELQHTNRLN	1004
990	Db	990	IHIPOYTEKNPLMYLTGVQTDKAGDISCNADINPLKIGQTSSSVSFKSENFRHTKELN	1049
1005	Qy	1005	GSNTQCQVVRCHLQGLAKGTEVSGVLLRLVHNSFFRRRAKFKSTVTVSTFELGTEGSLQ	1064
1050	Db	1050	CRTASCNVTCMLKDVHMKGEPVNVVTRINWGTTFASSTPQTQVLTAAAEINTYNPEIY-	1108
1065	Qy	1065	LTEASRWSESLLEVQTRILISLMI-----LIGSVLGGLLLLALLVFLCLWK	1111
1109	Db	1109	-----VIEDNTVTIPLMIKPKDEKAEAVTPTGVIGSIIAGILLALLVALWLK	1155
1112	Qy	1112	LGFFAHK-----KIPES	1123
1156	Db	1156	LGFFKRYEKWTKNPDE	1172

RESULT 5

US-09-949-016-6189

; Sequence 6189, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949, 016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241, 755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237, 768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231, 498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 0.766641 Seconds  
(without alignments)  
2761.097 Million cell updates/sec

Title: US-09-647-544-7  
Perfect score: 114  
Sequence: 1 KLGFFAHKKIPEEKREKLEQ 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:.\*  
1: PIR1:.\*  
2: PIR2:.\*  
3: PIR3:.\*  
4: PIR4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	46.5	214	2 A11073	probable membrane
2	51	44.7	672	2 G69503	signal-transducing
3	51	44.7	856	2 H64552	endopeptidase Clp
4	51	44.7	1045	2 S60571	integrin alpha v c
5	50	43.9	357	2 AC1104	B. subtilis YacL p
6	50	43.9	505	1 S77034	protein kinase pkn
7	50	43.9	1034	2 A36108	integrin alpha-v c
8	50	43.9	1044	2 T10050	integrin alpha-v c
9	50	43.9	1048	2 A27421	integrin alpha-5 c
10	49.5	43.4	236	2 T12766	probable lipoprote
11	49.5	43.4	959	1 B60017	outer capsid prote
12	49	43.0	357	2 AC1466	B. subtilis YacL p
13	49	43.0	615	2 T06108	hypothetical prote
14	49	43.0	853	2 S74279	hypothetical prote
15	49	43.0	856	2 C71956	probable endopepti
16	48.5	42.5	407	2 C273325	M2 protein precurs
17	48	42.1	421	2 H90433	hypothetical prote
18	48	42.1	573	2 J55767	squalene monooxyge
19	48	42.1	962	2 JC5808	G protein-coupled
20	48	42.1	1464	2 T13716	bazooka gene prote
21	48	42.1	1526	2 JN0598	DNA topoisomerase
22	47	41.2	102	2 E75077	hypothetical prote
23	47	41.2	112	2 E70433	flagellar switch p
24	47	41.2	206	2 T16946	hypothetical prote
25	47	41.2	512	2 AC3203	IS3 family transpo
26	47	41.2	512	2 AD2835	IS3 family transpo
27	47	41.2	512	2 AD3049	IS3 family transpo
28	47	41.2	512	2 G97622	probable transposa
29	47	41.2	512	2 G98236	probable transposa

30	47	41.2	692	2 S37976	hypothetical prote
31	47	41.2	851	2 T02267	trehalose-6-phosph
32	46	40.4	278	2 D84492	hypothetical prote
33	46	40.4	308	2 S11153	oligopeptide trans
34	46	40.4	308	2 E95220	hypothetical prote
35	46	40.4	323	2 E98084	hypothetical prote
36	46	40.4	450	2 S37900	hypothetical prote
37	46	40.4	544	2 AH2971	hypothetical prote
38	46	40.4	544	2 E98311	probable ATP-bindi
39	46	40.4	847	2 C96703	hypothetical prote
40	46	40.4	1037	2 A60163	glycoprotein Iib -
41	46	40.4	1407	1 S28589	trichohyalin - rab
42	45.5	39.9	570	2 F70332	proline-tRNA synth
43	45	39.5	273	2 T47612	hypothetical prote
44	45	39.5	325	2 T18283	hypothetical prote
45	45	39.5	333	2 T08850	alternative respir

ALIGNMENTS

RESULT 1

A11073  
probable membrane protein smp [imported] - Salmonella enterica subsp. enterica serovar Typhimurium  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: A11073  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: A11073  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-214 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD03408.1; PID:gl6505677; GSPDB:GN00176  
C:Genetics:  
A:Gene: smp

Query Match 46.5%; Score 53; DB 2; Length 214;  
Best Local Similarity 71.4%; Pred. No. 5.6;  
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Qy 6 AHKKIPEEKREK 19  
Db 200 ASRPVPEEREK 213

RESULT 2

G69503  
signal-transducing histidine kinase homolog - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C:Accession: G69503  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Nature 390, 364-370, 1997  
A:Authors: Uitterlinden, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.; Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo.  
A:Reference number: A69503; MUID:98049343; PMID:9389475  
A:Accession: G69503  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-672 <KLE>  
A:Cross-references: UNIPROT:O28247; GB:AE000963; GB:AE000782; NID:g2689286; PIDN:AAB8922

Query Match 44.7%; Score 51; DB 2; Length 672;

```

Best Local Similarity 55.6%; Pred. No. 33;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 FAHKKIPPEEKREKLEQ 22
Db 572 FAFKMEDEERRELLKQ 589

RESULT 3
H64552
endopeptidase Clp ATP-binding chain B - Helicobacter pylori (strain 26695)
N/Alternate names: ATP-dependent Clp proteinase regulatory chain
N/Contains: adenosinetriphosphatase (EC 3.6.1.3)
C/Species: Helicobacter pylori
C/Date: 09-Aug-1997 #sequence_revision 15-Aug-1997 #text_change 09-Jul-2004
C/Accession: H64552
R/Tomb, J.F.; White, O.; Karlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalaf, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A/Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A/Reference number: A64520; MUID: 97394467; PMID: 9252185
A/Accession: H64552
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-856 <TOM>
A/Cross-references: UNIPROT:P71404; GB:AE000545; GB:AE000511; NID:g2313349; PIDN:AAD0733
C/Function:
A/Description: allows clpP to hydrolyze polypeptides and proteins, probably by a chaperon
e activity; ATP hydrolysis is required for Clp hydrolysis of proteins but not of smaller
C/Superfamily: endopeptidase Clp ATP-binding chain
C/Keywords: ATP; duplication; hydrolase; molecular chaperone; nucleotide binding; P-loop
P/204-211/Region: nucleotide-binding motif A (P-loop)
F/271-276/Region: nucleotide-binding motif B
F/605-612/Region: nucleotide-binding motif A (P-loop)
F/673-678/Region: nucleotide-binding motif B
F/210/Binding site: ATP (Lys) #status predicted
F/611/Binding site: ATP (Lys) #status predicted

Query Match 44.7%; Score 51; DB 2; Length 856;
Best Local Similarity 45.0%; Pred. No. 42;
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 GFFAHKKIPPEEKREKLEQ 22
Db 496 GEIYSKIPENKKKEELQR 515

RESULT 4
S60571
integrin alpha v chain precursor - Iberian ribbed newt
C/Species: Pleurodeles waltliffi (Iberian ribbed newt)
C/Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C/Accession: S60571
R/Alifanari, D.; Whittaker, C.A.; Desimone, D.W.; Darribere, T.
Dev. Biol. 170, 249-261, 1995
A/Title: Integrin alpha-(v) subunit is expressed on mesodermal cell surfaces during amph
A/Reference number: S60571; MUID: 95377519; PMID: 7649360
A/Accession: S60571
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-1045 <ALF>
A/Cross-references: UNIPROT:Q91292; EMBL:X81108; NID:g1008137; PIDN:CAA57014.1; PID:g100
C/Superfamily: integrin alpha-2b chain
F/1-29/Domain: signal sequence #status predicted <SIG>
F/30-1045/Product: integrin alpha v chain #status predicted <MAT>

Query Match 44.7%; Score 51; DB 2; Length 1045;
Best Local Similarity 42.9%; Pred. No. 51;
Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 KLGFPAHKKIPPEEKREKLE 21

```

```

Db 1014 KFGFFKRVPPQBEQRLEQLQ 1034

RESULT 5
AC1104
B. subtilis YacL protein homolog lmo0234 [imported] - Listeria monocytogenes (strain EGD-
C/Species: Listeria monocytogenes
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C/Accession: AC1104
R/Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Barche, P.; Bloecker,
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussutget, O.; Entian, K.D.; Feihl, H.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A/Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat
ok, C.; Schluster, T.; Simoes, N.; Tlierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A/Title: Comparative Genomics of Listeria species.
A/Reference number: AB1077; MUID: 21537279; PMID: 11679669
A/Accession: AC1104
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-357 <GLA>
A/Cross-references: UNIPROT:Q48762; GB:NC_003210; PIDN:CAD00761.1; PID:g16409599; GSPDB:C
A/Experimental source: strain EGD-e
C/Genetics:
A/Gene: lmo0234
C/Superfamily: conserved hypothetical protein yacL

Query Match 43.9%; Score 50; DB 2; Length 357;
Best Local Similarity 71.4%; Pred. No. 24;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 AHKKIPPEEKREK 19
Db 146 AKRTPPEEKTEK 159

RESULT 6
S77034
protein kinase pKna (EC 2.7.1.-), 55K - Synecocystis sp. (strain PCC 6803)
N/Alternate names: protein sli10776
C/Species: Synecocystis sp.
A/Variety: PCC 6803
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: S77034
R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shampo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
s.

A/Reference number: S74322; MUID: 97061201; PMID: 8905231
A/Accession: S77034
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-505 <KAN>
A/Cross-references: UNIPROT:P54735; EMBL:D64005; GB:AB001339; NID:g1001779; PIDN:BAAL0724
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C/Genetics:
A/Gene: pKna
C/Superfamily: Synecocystis sp. protein kinase pKna, 55K; protein kinase homology
C/Keywords: phosphotransferase; protein kinase
F/7-268/Domain: protein kinase homology <KIN>

Query Match 43.9%; Score 50; DB 1; Length 505;
Best Local Similarity 58.8%; Pred. No. 34;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 6 AHKKIPPEEKREKLEQ 22
Db 376 AEQKIENKQRELEQ 392

RESULT 7

```

A;Reference number: S47541; MUID:94368864; PMID:7522056  
A;Accession: S47541  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-61 <DOW>  
A;Cross-references: EMBL:U07375  
R;Fitzgerald, L.A.; Poncz, M.; Steiner, B.; Rall Jr., S.C.; Bennett, J.S.; Phillips, D.R.  
Biochemistry 26, 8158-8165, 1987  
A;Title: Comparison of cDNA-derived protein sequences of the human fibronectin and vitro-  
A;Reference number: A90526; MUID:88163472; PMID:2450560  
A;Accession: B29418  
A;Molecule type: mRNA  
A;Residues: 1-433 <FIT>  
R;Suzuki, S.; Agraves, W.S.; Pytela, R.; Arai, H.; Krusius, T.; Pierschbacher, M.D.; Ru-  
Proc. Natl. Acad. Sci. U.S.A. 83, 8614-8618, 1986  
A;Title: cDNA and amino acid sequences of the cell adhesion protein receptor recognizing  
A;Reference number: A26482; MUID:87041504; PMID:2430295  
A;Accession: A26482  
A;Molecule type: mRNA  
A;Residues: 413-1048 <SU2>  
R;Cheresh, D.A.; Smith, J.W.; Cooper, H.M.; Quaranta, V.  
Cell 57, 59-69, 1989  
A;Title: A novel vitronectin receptor integrin (alpha-v beta-x) is responsible for distri-  
A;Reference number: A32287; MUID:89195223; PMID:2467745  
A;Accession: A32287  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 31-35,'X','37'-41 <CHE>  
A;Experimental source: melanoma cell M21  
A;Accession: B32287  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 31-33,'X','35','X','37'-41 <CH2>  
A;Experimental source: lung carcinoma cell UCLA-P3  
R;Lim, S.C.T.; Plow, E.F.; D'Souza, S.E.; Cheresh, D.A.; Frelinger III, A.L.; Ginsberg, I.  
J. Biol. Chem. 264, 3742-3749, 1989  
A;Title: Isolation and characterization of a platelet membrane protein related to the vi-  
A;Reference number: A30298; MUID:89139425; PMID:2465293  
A;Accession: A30298  
A;Molecule type: protein  
A;Residues: 31-35,'X','37'-40 <LAM>  
R;Smith, J.W.; Cheresh, D.A.  
J. Biol. Chem. 265, 2168-2172, 1990  
A;Title: Integrin (alpha-vbeta-3)-ligand interaction. Identification of a heterodimeric  
A;Reference number: A35035; MUID:90130470; PMID:1688848  
A;Accession: A35035  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 66-72;169-171,'X',173-176;221-230;255-258,'X',260;325-328;342-351;466-473 <S<  
C;Genetics:  
A;Gene: GDB:ITGAV; VNRA  
A;Cross-references: GDB:L20491; OMIM:193210  
A;Map position: 2q31-2q32  
C;Superfamily: integrin alpha-2b chain  
C;Keywords: cell adhesion; duplication; glycoprotein; heterodimer; receptor; transmembra-  
P;1-30/Domain: signal sequence #status predicted <SIG>  
P;31-1048/Product: integrin alpha-v chain #status experimental <MAT>  
P;31-992/Domain: extracellular #status predicted <EXT>  
F;993-1016/Domain: transmembrane #status predicted <TMN>  
F;1017-1048/Domain: intracellular #status predicted <INT>

Query Match 43.9%; Score 50; DB 2; Length 1048;  
Best Local Similarity 38.1%; Pred. No. 70;  
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 1 KLGFPAHKKIPDEEKREKLE 21  
::||| : ||| : |:  
Db 1017 RMGFKKVRPPQEERQLQ 1037

RESULT 10  
T12766  
probable lipoprotein yokB - Bacillus subtilis phage SPBC2



C;Species: *Bacillus subtilis* phage SPBc2  
 C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 09-Jul-2004  
 C;Accession: T12766; F69907  
 R;Lazarevic, V.; Duesterhoeft, A.; Solido, B.; Hilbert, H.; Mael, C.; Karamata, D.  
 submitted to the EMBL Data Library, August 1997  
 A;Description: The complete nucleotide sequence of the *Bacillus subtilis* SPBc2 prophage  
 A;Reference number: 217583  
 A;Accession: T12766  
 A;Status: translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-236 <LAZ>  
 A;Cross-references: UNIPROT:O64016; EMBL:AF020713; NID:G3025478; PID:G3025480; PIDN:AACT  
 R;Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd  
 A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen  
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
 A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
 A;Authors: Schleicher, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
 akouchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
 A;Authors: Yoshikawa, H.P.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A;Reference number: A69580; MUID:98044033; PMID:9384377  
 A;Accession: F69907  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-236 <KUN>  
 A;Cross-references: GB:299115; GB:AL009126; NID:G2634478; PIDN:CAB14083.1; PID:e1183612;  
 A;Experimental source: strain 168  
 C;Genetics:  
 A;Gene: yokB

Query Match 43.4%; Score 49.5; DB 2; Length 236;  
 Best Local Similarity 84.6%; Pred. No. 19;  
 Matches 11; Conservative 1; Mismatches 0; Indels 1; Gaps 1;  
 Qy 8 KKI-PEEKREK 19  
 ||| |||||  
 Db 205 KKVSPPEEKREK 217

RESULT 11  
 B60017  
 outer capsid protein VP2 - bluetongue virus (serotype 3, strain South Africa-VACC)  
 C;Species: bluetongue virus  
 C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
 C;Accession: B60017  
 R;Gould, A.R.; Pritchard, L.I.  
 Virus Res. 17, 31-52, 1990  
 A;Title: Relationships amongst bluetongue viruses revealed by comparisons of capsid and  
 A;Reference number: A60017; MUID:91021485; PMID:2171239  
 A;Accession: B60017  
 A;Molecule type: genomic RNA  
 A;Residues: 1-959 <GOU>  
 A;Cross-references: UNIPROT:Q06998; GB:X55801; NID:G297130; PIDN:CAA39323.1; PID:G297132  
 C;Genetics:  
 A;Map position: segment 2  
 C;Superfamily: bluetongue virus VP2 protein  
 C;Keywords: capsid protein; glycoprotein  
 F;749,910/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 Query Match 43.4%; Score 49.5; DB 1; Length 959;  
 Best Local Similarity 57.1%; Pred. No. 75;  
 Matches 12; Conservative 1; Mismatches 3; Indels 5; Gaps 1;  
 Qy 2 LGFFAHKKIPPEEKREKLEQ 22  
 ||| |||||  
 Db 791 LNFF-----PSYKREKLE 806

## RESULT 12

AC1466  
 B. subtilis yacL protein homolog lin0266 [imported] - *Listeria innocua* (strain Clip11262)  
 C;Species: *Listeria innocua*  
 C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
 C;Accession: AC1466  
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,  
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussauguet, O.; Entian, K.D.; Faihi, H.;  
 D.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat  
 ok, C.; Schluteter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
 C.; Title: Comparative genomics of *Listeria* species.  
 A;Reference number: AB1077; MUID:21537279; PMID:11679669  
 A;Accession: AC1466  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-357 <GLA>  
 A;Cross-references: UNIPROT:Q92F41; GB:AL592022; PIDN:CAC95499.1; PID:gl6412695; GSPDB:G  
 A;Experimental source: strain Clip11262  
 C;Genetics:  
 A;Gene: lin0266  
 C;Superfamily: conserved hypothetical protein yacL

Query Match 43.0%; Score 49; DB 2; Length 357;  
 Best Local Similarity 71.4%; Pred. No. 33;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 6 AHKKIPPEEKREK 19  
 ||| |||||  
 Db 146 AKKTPPEEKREK 159

## RESULT 13

T06108  
 hypothetical protein TSJ17.190 - *Arabidopsis thaliana*  
 C;Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C;Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 09-Jul-2004  
 C;Accession: T06108  
 R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.  
 submitted to the Protein Sequence Database, March 1999  
 A;Reference number: Z15184  
 A;Accession: T06108  
 A;Molecule type: DNA  
 A;Residues: 1-615 <BEV>  
 A;Cross-references: UNIPROT:Q9SMQ3; EMBL:AL035708; GSPDB:GN00062; ATSP:TSJ17.190  
 A;Experimental source: cultivar Columbia; BAC clone TSJ17  
 C;Genetics:  
 A;Gene: ATSP:TSJ17.190  
 A;Map position: 4

Query Match 43.0%; Score 49; DB 2; Length 615;  
 Best Local Similarity 71.4%; Pred. No. 56;  
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 8 KKIPPEEKREKLE 21  
 ||| |||||  
 Db 396 KKIEEKKEEKLE 409

## RESULT 14

S74279  
 hypothetical protein YCJ061c - yeast (*Saccharomyces cerevisiae*)  
 N;Alternate names: hypothetical protein YCJ060c  
 C;Species: *Saccharomyces cerevisiae*  
 C;Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
 C;Accession: S74279; S19392; S19391; S29373; S21360  
 R;Voet, M.; Volckaert, G.  
 submitted to the Protein Sequence Database, September 1996  
 A;Reference number: S74277  
 A;Accession: S74279



A:Molecule type: DNA  
A:Residues: 1-853 <VOE>  
A:Cross-references: UNIPROT:P25588; EMBL:X59720; NID:g1907116; PIDN:CAA42405.1; PID:e309  
A:Note: this is a revision to the sequence from reference S19391  
R:Rasmussen, S.W.; von Wettstein, D.  
submitted to the Protein Sequence Database, March 1992  
A:Reference number: S19391  
A:Accession: S19392  
A:Molecule type: DNA  
A:Residues: 'MKLPIMHLVVMYIVNPIVKKRMKML', 346-619, 'LP', 621, 'P', 623-638, 'LRTIALM' <RAS>  
A:Cross-references: EMBL:X59720; MIPS:YCL061c  
A:Note: this sequence has been revised in reference S74279  
A:Note: this was assumed to be the complete sequence of protein YCL061c  
A:Accession: S19391  
A:Molecule type: DNA  
A:Residues: 1-314, 'LEW' <RAW>  
A:Cross-references: EMBL:X59720  
A:Note: this sequence has been revised in reference S74279  
A:Note: this was assumed to be protein YCL060c  
R:Kern, L.  
Nucleic Acids Res. 18, 5279, 1990  
A:Title: The URK1 gene of Saccharomyces cerevisiae encoding uridine kinase.  
A:Reference number: S29373; MUID:90384830; PMID:2169608  
A:Accession: S29373  
A:Molecule type: DNA  
A:Residues: 417-504, 'V', 506-564, 566-596 <KER>  
A:Cross-references: EMBL:X53998; NID:g4771; PIDN:CAA37945.1; PID:g4772  
C:Genetics:  
A:Map position: 3L  
A:Note: YCL061c

Query Match 43.0%; Score 49; DB 2; Length 853;

Best Local Similarity 56.2%; Pred. No. 78;

Matches 9; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 7 HKKIPBEEKREKLEQ 22

DB 277 NKRIQKREKREKLE 292

#### RESULT 15

C71956

probable endopeptidase Clp ATP-binding chain - Helicobacter pylori (strain J99)

N:Alternate names: ATP-dependent Clp proteinase regulatory chain

N:Contains: adenosinetriphosphatase (EC 3.6.1.3)

C:Species: Helicobacter pylori

A:Variety: strain J99

C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004

C:Accession: C71956

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;

Ives, C.; Gibson, R.; Merberg, D.; Millis, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path

A:Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: C71956

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-856 <ARN>

A:Cross-references: UNIPROT:Q92MH1; GB:AE001462; GB:AE001439; NID:g4154760; PIDN:AAD0582

A:Experimental source: strain J99

C:Genetics:

A:Gene: clpB

C:Function:

A:Description: allows clpP to hydrolyze polypeptides and proteins, probably by a chaperon

e activity; ATP hydrolysis is required for Clp hydrolysis of proteins but not of smaller

C:Superfamily: endopeptidase Clp ATP-binding chain

C:Keywords: ATP; hydrolase; molecular chaperone; nucleotide binding; P-loop

F:204-211/Region: nucleotide-binding motif A (P-loop)

F:271-276/Region: nucleotide-binding motif B

F:605-612/Region: nucleotide-binding motif A (P-loop)

F:673-678/Region: nucleotide-binding motif B

F:210/Binding site: ATP (Lys) #status predicted

F:611/Binding site: ATP (Lys) #status predicted

Query Match 43.0%; Score 49; DB 2; Length 856;  
Best Local Similarity 53.3%; Pred. No. 78;  
Matches 8; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 7 HKKIPBEEKREKLE 21

DB 500 YSKIPEKEKEKELQ 514

Search completed: April 6, 2005, 12:17:14

Job time: 1.76664 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 3.25607 Seconds  
(without alignments)  
3459.921 Million cell updates/sec

Title: US-09-647-544-7  
Perfect score: 114  
Sequence: 1 KLGFPAHKIPBEKREKLEQ 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114	100.0	1167	1 ITAG_HUMAN	O75578 homo sapien
2	88	77.2	288	2 QBEM12	Q8Bm12 mus musculus
3	55	48.2	997	2 Q7PXX0	Q7Pxx0 anopheles g
4	53	46.5	214	1 SMP_SALTI	Q820t9 salmonella
5	53	46.5	214	1 SMP_SALTY	Q82jv0 salmonella
6	53	46.5	492	2 Q73FL3	Q73fl3 wolbachia p
7	53	46.5	1033	2 Q42598	Q42598 xenopus lae
8	52	45.6	1021	2 Q9GSF4	Q9gsf4 podocoryne
9	51	44.7	128	1 YTE1_BACAN	Q9rmz2 bacillus an
10	51	44.7	134	2 Q61ED6	Q61ed6 ovis aries
11	51	44.7	180	2 Q8TJ36	Q8tzj6 pyrococcus
12	51	44.7	672	2 Q28247	Q28247 archaeoglob
13	51	44.7	856	1 CLPB_HELPY	P71404 helicobacte
14	51	44.7	1013	2 Q8BLA0	Q8blao mus musculus
15	51	44.7	1045	2 Q91292	Q91292 pleurodeles
16	51	44.7	1914	2 Q68FD0	Q68fd0 mus musculus
17	51	44.7	1957	2 Q91YC9	Q91yc9 mus musculus
18	50	43.9	145	2 Q8X3V9	Q6x3v9 bacillus th
19	50	43.9	145	2 Q7WSF9	Q7wsf9 bacillus th
20	50	43.9	175	2 Q80Y67	Q80y67 mus musculus
21	50	43.9	268	2 Q61MW4	Q61mw4 methanococc
22	50	43.9	357	1 Y234_LISMO	Q48762 listeria mo
23	50	43.9	357	2 Q724H8	Q724h8 listeria mo
24	50	43.9	403	2 Q8XYF8	Q8xyf8 trypanosoma
25	50	43.9	505	1 SPKD_SINY3	P54735 synecocyst
26	50	43.9	1034	1 ITAV_CHICK	P26008 gallus gall
27	50	43.9	1044	1 ITAV_MOUSE	P43406 mus musculus
28	50	43.9	1047	1 ITAV_BOVIN	P80746 bos taurus
29	50	43.9	1048	1 ITAV_HUMAN	P06756 homo sapien
30	49.5	43.4	236	2 Q840I6	Q640i6 bacterioph
31	49.5	43.4	236	2 Q32005	O32005 bacillus su

32	49.5	43.4	635	2	Q96LR8	Q96lr8 homo sapien
33	49.5	43.4	846	2	Q86W98	Q86w98 homo sapien
34	49.5	43.4	959	1	VP2_BT3V3V	Q6998 bluetongue
35	49.5	43.4	959	2	Q6RYD5	Q6ryd5 bluetongue
36	49.5	43.4	1372	2	Q9HAW1	Q9haw1 homo sapien
37	49.5	43.4	1388	2	Q9HAW2	Q9haw2 homo sapien
38	49.5	43.4	1985	2	Q7RC59	Q7rc59 plasmodium
39	49.5	43.4	2187	2	Q9H197	Q9h197 homo sapien
40	49.5	43.4	2254	2	Q9HCY0	Q9hcy0 homo sapien
41	49.5	43.4	2627	2	Q9C0H4	Q9c0h4 homo sapien
42	49	43.0	357	1	Y266_LISIN	Q92f41 listeria in
43	49	43.0	615	2	Q9SMQ3	Q9smq3 arabidopsis
44	49	43.0	856	1	CLPB_HELPJ	Q92mh1 helicobacte
45	49	43.0	1096	1	MRC1_YEAST	P25588 saccharomyc

#### ALIGNMENTS

#### RESULT 1

ID ITAG\_HUMAN STANDARD; PRT; 1167 AA.  
AC O75578; Q9UH28;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Integrin alpha-10 precursor.  
GN Name=ITGAL0;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Articular chondrocytes;  
RX MEDLINE=98352078; PubMed=9685391; DOI=10.1074/jbc.273.32.20383;  
RA Camper L., Hellman U., Lundgren-Akerlund E.;  
RT "Isolation, cloning, and sequence analysis of the integrin subunit  
alpha10, a beta1-associated collagen binding integrin expressed on  
chondrocytes.";  
RL J. Biol. Chem. 273:20383-20389(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Endothelial cells, and Heart;  
RX MEDLINE=20169197; PubMed=10702680;  
RA Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,  
Wang S.-X., Langley R., Krissansen G.W.;  
RT "The integrin alpha10 subunit: expression pattern, partial gene  
structure, and chromosomal localization.";  
RL Cytogenet. Cell Genet. 87:238-244(1999).  
CC -!- FUNCTION: Integrin alpha-10/beta-1 is a receptor for collagen.  
CC -!- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-10  
CC associates with beta-1.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- TISSUE SPECIFICITY: Widely expressed with highest expression in  
CC muscle and heart. Found in articular cartilage.  
CC -!- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins  
CC with I-domains do not undergo protease cleavage.  
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.  
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.  
CC -!- SIMILARITY: Contains 1 VWFA domain.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; AF074015; AAC31952.1; -  
EMBL; AF112345; AAF21944.1; -  
EMBL; AF172723; AAF61638.1; -



DR GO: GO:0007160; P:cell-matrix adhesion; IEA.  
 DR GO: GO:0007229; P:integrin-mediated signaling pathway; IEA.  
 DR InterPro: IPR000413; Integrin\_alpha.  
 DR Pfam: PF00357; Integrin\_alpha; 1.  
 DR PROSITE: PS00242; INTEGRIN\_ALPHA; UNKNOWN\_1.  
 KW Cell adhesion; Integrin; Transmembrane.  
 FT NON\_TER 1  
 SQ SEQUENCE 288 AA; 31949 MW; 3F6200F9C9475E85 CRC64;  
 Query Match 77.2%; Score 88; DB 2; Length 288;  
 Best Local Similarity 77.3%; Pred. No. 0.00072;  
 Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 KLGFPAHKIPPEEKREKLEQ 22  
 Db 267 KLGFTRKIPKEESEKLEQ 288  
 RESULT 3  
 ID Q7PXX0 PRELIMINARY; PRT; 997 AA.  
 AC Q7PXX0;  
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
 DE AGCP12053 (Fragment)  
 GN Name=agCS6510; ORFNames=ENSANGG00000014880;  
 OS Anopheles gambiae str. PEST.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoides; Anopheles.  
 OX NCBI\_TaxID=180454;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PEST;  
 RA Anopheles Genome Sequencing Consortium;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AAA01008987; EAA00964.1; -  
 DR InterPro: IPR007110; IG-like.  
 DR PROSITE: PS50835; IG-like; 6.  
 FT NON\_TER 1  
 FT NON\_TER 997 997  
 SQ SEQUENCE 997 AA; 113756 MW; 3C3110C77B646F94 CRC64;  
 Query Match 48.2%; Score 55; DB 2; Length 997;  
 Best Local Similarity 71.4%; Pred. No. 83;  
 Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 Qy 8 KKIPEEKREKLE 21  
 Db 840 KKVEEKKEKVE 853  
 RESULT 4  
 ID - SMP SALTI STANDARD; PRT; 214 AA.  
 AC Q8Z0T9;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Protein smp precursor.  
 GN Name=smp; OrderedLocusNames=STY4924, t4616;  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RA MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;  
 RX Parkhill J., Dougan G., James K.D., Thomson N.K., Pickard D., Wain J.,

RA Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
 Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,  
 Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
 Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
 Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,  
 Whitehead S., Barrell B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 enterica serovar Typhi CT18.";  
 RL Nature 413:848-852(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ty2 / ATCC 700931;  
 RX MEDLINE=22531367; PubMed=12644504;  
 RX DOI=10.1128/JB.185.7.2330-2337.2003;  
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
 Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;  
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
 and CT18";  
 RL J. Bacteriol. 185:2330-2337(2003).  
 CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).  
 CC -1- SIMILARITY: Belongs to the smp family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; AL627284; CAD03408.1; -  
 DR EMBL; AE016849; AA072048.1; -  
 KW Complete proteome; Membrane; Signal.  
 FT SIGNAL 1 30 Potential.  
 FT CHAIN 31 214 Protein smp.  
 SQ SEQUENCE 214 AA; 24279 MW; 2683D0DF9174BF80 CRC64;  
 Query Match 46.5%; Score 53; DB 1; Length 214;  
 Best Local Similarity 71.4%; Pred. No. 33;  
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Qy 6 AHKKIPEEKREK 19  
 Db 200 ASKVPPEEREK 213  
 RESULT 5  
 ID - SMP SALTY STANDARD; PRT; 214 AA.  
 AC Q8ZJVO;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Protein smp precursor.  
 GN Name=smp; OrderedLocusNames=STM4577;  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Lt2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;  
 RX McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 LT2.";  
 RL Nature 413:852-856(2001).  
 CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).

DT	01-JAN-1998	(TrEMBLrel. 05, Last sequence update)
DE	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)
DE	Integrin alphav subunit.	
OS	Xenopus laevis	(African clawed frog).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;	
OC	Xenopodinae; Xenopus.	
OX	NCBI_TaxID=8355;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE-98211587; PubMed-9551862;	
RA	Joois T.O., Reintsch W.E., Brinker A., Klein C., Hausen P.;	
RT	"Cloning of the Xenopus integrin alpha(v) subunit and analysis of its	
RL	distribution during early development."	
RL	Int. J. Dev. Biol. 42:171-179(1998).	
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (by similarity).	
CC	-1- SIMILARITY: Belongs to the integrin alpha chain family.	
DR	ENBL; U92006; AABG2090.1; -.	
DR	HSSP; P06756; 1L5G.	
DR	GO; GO:0016021; C:integral to membrane; IEA.	
DR	GO; GO:0008305; C:integrin complex; IEA.	
DR	GO; GO:0005515; F:protein binding; IEA.	
DR	GO; GO:0007160; P:cell-matrix adhesion; IEA.	
DR	GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.	
DR	InterPro; IPR000413; Integrin_alpha.	
DR	Pfam; PF00357; Integrin_alpha; 1.	
DR	PRINTS; PR01185; INTEGRIN_A.	
DR	SMART; SW00191; Int_alpha; 5.	
DR	PROSITE; PS00242; INTEGRIN_ALPHA; 1.	
KW	Cell adhesion; Integrin; Transmembrane.	
SQ	SEQUENCE 1033 AA; 114024 MW; 35F73B0765C06BD6 CRC64;	
	Query Match 46.5%; Score 53; DB 2; Length 1033;	
	Best Local Similarity 47.8%; Pred.No. 1.6e+02;	
	Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;	
QY	1 KLGFFAHKKIPPEEKREKLE 21	
	:      :	
DB	1002 KLGFPRVVRPPQEETEREQLQ 1022	
RESULT 8		
ID	Q9GSF4 PRELIMINARY; PRT; 1021 AA.	
QC	Q9GSF4;	
AC	01-MAR-2001 (TrEMBLrel. 16, Created)	
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)	
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)	
DE	Integrin alpha chain.	
GN	Name=Inta;	
OS	Podocoryne carnea.	
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;	
OC	Hydractiniidae; Podocoryne.	
OX	NCBI_TaxID=6096;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE-21375962; PubMed-11482899; DOI=10.1006/cbir.2000.0708;	
RA	Reber-Muller S., Scuder R., Muller P., Yanze N., Schmid V.;	
RT	"Integrin and talin in the jellyfish Podocoryne carnea.";	
RL	Cell Biol. Int. 25:753-769(2001).	
RL	[2]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE-21375962; PubMed-11482899; DOI=10.1006/cbir.2000.0708;	
RA	Reber-Muller S., Scuder R., Muller P., Yanze N., Schmid V.;	
RT	"Integrin and talin in the jellyfish Podocoryne carnea.";	
RL	Cell Biol. Int. 25:753-769(2001).	
RL	[2]	
RP	SEQUENCE FROM N.A.	
RX	Reber-Muller S., Studer R., Mueller P., Yanze N., Schmid V.;	
RL	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.	
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (by similarity).	
CC	-1- SIMILARITY: Belongs to the integrin alpha chain family.	
DR	ENBL; AF308651; AG25993.1; -.	
DR	GO; GO:0016021; C:integral to membrane; IEA.	
DR	GO; GO:0008305; C:integrin complex; IEA.	
DR	GO; GO:0005515; F:protein binding; IEA.	
DR	GO; GO:0007160; P:cell-matrix adhesion; IEA.	
DR	GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.	
DR	InterPro; IPR000413; Integrin_alpha.	



```

RESULT 12
O28247 PRELIMINARY; PRT; 672 AA.
AC O28247;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Signal-transducing histidine kinase, putative.
GN OrderedLocusNames=AF2032;
OS Archaeoglobus fulgidus.
OS Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475; DOI=10.1038/37052;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M.L., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Karlavage A.R., Graham M.H., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.R.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.J.,
RA Peterson S.N., Reich C.I., McNeil L.K., Badger J.H., Glodek A.,
RA Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L.A.,
RA Uterback T.R., Cotton M.D., Spriggs T., Artisch P., Kaine B.P.,
RA Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C.,
RA Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O.,
RA Woese C.R., Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AE000963; AB89224.1; -.
DR PIR; G69503; G69503.
DR TIGR; AF2032; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR InterPro; IPR003661; His_kin_N.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS.
DR InterPro; IPR000700; PAS-assoc_C.
DR Pfam; PF00512; HisKA; 1.
DR Pfam; PF00785; PAC; 3.
DR Pfam; PF00989; PAS; 3.
DR SMART; SM00388; HisKA; 1.
DR SMART; SM00086; PAC; 3.
DR SMART; SM00091; PAS; 4.
DR TIGRFAMs; TIGR00229; sensory_box; 4.
DR PROSITE; PS50113; PAC; 3.
DR PROSITE; PS50112; PAS; 3.
KW Complete proteome; Kinase.
SQ SEQUENCE 672 AA; 78620 MW; F95DC675BBD8F749 CRC64;

Query Match 44.7%; Score 51; DB 2; Length 672;
Best Local Similarity 55.6%; Pred. No. 2e+02;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 5 FAHKKIPEEKREKLEQ 22
Db 572 FAPKSMEDERRELLKQ 589
||| : :||| :|
| : :| :| :|

RESULT 13
CLPB_HELPY STANDARD; PRT; 856 AA.
AC P71404;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Chaperone clpB.

```

Name=clpB; OrderedLocusNames=HP0264;  
 Helicobacter pylori (Campylobacter pylori).  
 Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
 OC Helicobacteraceae; Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCTC 11637;  
 RX PubMed=9440536;  
 RA Allan E., Mullany P., Tabaqchali S.;  
 RT "Construction and characterization of a Helicobacter pylori clpB  
 RT mutant and role of the gene in the stress response.";  
 RL J. Bacteriol. 180:426-429(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=26695 / ATCC 700392;  
 RX MEDLINE=97394467; PubMed=9252185; DOI=10.1038/41483;  
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,  
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S.R.,  
 RA Dougherty B.A., Nelson K.E., Quackenbush J., Zhou L., Kirkness E.F.,  
 RA Peterson S.N., Loftus B.J., Richardson D.L., Dodson R.J., Khalak H.G.,  
 RA Glodek A., McKenney K., FitzGerald L.M., Lee N., Adams M.D.,  
 RA Hickey E.K., Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D.,  
 RA Kelley J.M., Cotton M.D., Weidman J.F., Fujii C., Bowman C.,  
 RA Wathey L., Wallin E., Hayes W.S., Borodovsky M., Karp P.D.,  
 RA Smith H.O., Fraser C.M., Venter J.C.;  
 RT "The complete genome sequence of the gastric pathogen Helicobacter  
 RT pylori.";  
 RL Nature 388:539-547(1997).  
 CC -1- FUNCTION: Part of a stress-induced multi-chaperone system, it is  
 CC involved in the recovery of the cell from heat-induced damage, in  
 CC cooperation with dnaK, dnaJ and grpE. Acts before dnaK, in the  
 CC processing of protein aggregates. Protein binding stimulates the  
 CC ATPase activity; ATP hydrolysis unfolds the denatured protein  
 CC aggregates, which probably helps expose new hydrophobic binding  
 CC sites on the surface of clpB-bound aggregates, contributing to the  
 CC solubilization and refolding of denatured protein aggregates by  
 CC dnaK (By similarity). Necessary for surviving high-temperature  
 CC stress.  
 CC -1- SUBUNIT: Homohexamer. The oligomerization is ATP-dependent (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -1- DOMAIN: The N-terminal domain probably functions as a substrate-  
 CC discriminating domain, recruiting aggregated proteins to the clpB  
 CC hexamer and/or stabilizing bound proteins. The NBD2 domain is  
 CC responsible for oligomerization, whereas the NBD1 domain  
 CC stabilizes the hexamer probably in an ATP-dependent manner. The  
 CC movement of the coiled-coil domain is essential for clpB ability  
 CC to rescue proteins from an aggregated state, probably by pulling  
 CC apart large aggregated proteins, which are bound between the  
 CC coiled-coils motifs of adjacent clpB subunits in the functional  
 CC hexamer (By similarity).  
 CC -1- SIMILARITY: Belongs to the clpA/clpB family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk).  
 CC -----  
 CC EMBL; Y08238; CAA69406.1; -.  
 CC EMBL; AE000545; AAD07330.1; -.  
 CC PIR; H64552; H64552.  
 CC HSP; P03815; IJBK.  
 CC TIGR; HP0264; -.  
 CC InterPro; IPR003593; AAA\_ATPase.  
 CC InterPro; IPR003959; AAA\_ATPase\_cent.  
 CC InterPro; IPR001270; Chaprin\_clpA/B.  
 CC InterPro; IPR004176; Clp\_N.  
 CC Pfam; PF00004; AAA; 2.  
 CC Pfam; PF02861; Clp\_N; 2.



```

DR PRINTS; PRO0300; CLPPROTEASEA.
DR SMART; SMO0382; AAA; 2.
DR PROSITE; PS00870; CLPAB_1; 1.
DR PROSITE; PS00871; CLPAB_2; 1.
KW ATP-binding; Chaperone; Coiled coil; Complete proteome; Heat shock;
KW Repeat.
FT FT DOMAIN 1 141 N-terminal (By similarity).
FT FT DOMAIN 157 337 NBD1 (By similarity).
FT FT DOMAIN 338 545 Linker (By similarity).
FT FT DOMAIN 555 763 NBD2 (By similarity).
FT FT DOMAIN 764 856 C-terminal (By similarity).
FT FT DOMAIN 368 522 Coiled coil (By similarity).
FT FT NP_BIND 204 211 ATP 1 (By similarity).
FT FT NP_BIND 605 612 ATP 2 (By similarity).
FT FT CONFLICT 14 14 T -> A (in Ref. 1).
FT FT CONFLICT 33 33 Q -> L (in Ref. 1).
FT FT CONFLICT 59 59 Q -> E (in Ref. 1).
FT FT CONFLICT 62 62 R -> K (in Ref. 1).
FT FT CONFLICT 79 79 S -> N (in Ref. 1).
FT FT CONFLICT 103 103 R -> T (in Ref. 1).
FT FT CONFLICT 119 119 G -> S (in Ref. 1).
FT FT CONFLICT 131 131 A -> T (in Ref. 1).
FT FT CONFLICT 145 145 R -> A (in Ref. 1).
FT FT CONFLICT 149 149 D -> G (in Ref. 1).
FT FT CONFLICT 222 222 M -> V (in Ref. 1).
FT FT CONFLICT 435 435 A -> H (in Ref. 1).
FT FT CONFLICT 465 465 V -> A (in Ref. 1).
FT FT CONFLICT 643 644 MS -> IT (in Ref. 1).
FT FT CONFLICT 649 649 A -> P (in Ref. 1).
FT FT CONFLICT 734 734 E -> D (in Ref. 1).
FT FT CONFLICT 765 765 D -> G (in Ref. 1).
FT FT CONFLICT 836 836 V -> I (in Ref. 1).
FT FT CONFLICT 840 840 D -> G (in Ref. 1).
SQ SEQUENCE 856 AA; 96683 MW; E3902C7D989496AD CRC64;

Query Match 44.7%; Score 51; DB 1; Length 856;
Best Local Similarity 45.0%; Pred. No. 2.5e+02;
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 3 GFFAHHKPIPEEKKEEKLQ 22
Db 496 GEIYSKIPENKKEELQR 515

RESULT 14
Q8BLA0
ID Q8BLA0 PRELIMINARY; PRT; 1013 AA.
AC Q8BLA0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length
DE enriched library, clone:B230314P12 product:similar to TUMOR SUPPRESSOR
DE P53-BINDING PROTEIN 1 (P53-BINDING PROTEIN 1) (53BP1) (Fragment).
GN Name=Trp53bp1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).

RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanaki M.,
Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;
RA Adachi J., Aizawa K., Akimura T., Akakawa T., Bono H., Carninci P.,
Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
Hori F., Imotani K., Ishii Y., Itoh M., Kigawa I., Kasukawa T.,
Kurohara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
Nishi K., Nomura K., Numazaki R., Ohno N., Ohsato N., Okazaki Y.,
Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR MGD; MGI:1351320; Trp53bp1.
DR GO; GO:0000776; C:kinetochore; IDA.
DR GO; GO:0003684; F:damaged DNA binding; IDA.
DR GO; GO:0008134; F:transcription factor binding; TAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.
FT NON TER 1013 1013
SQ SEQUENCE 1013 AA; 109079 MW; 4AAE3FDA60E42F72 CRC64;

Query Match 44.7%; Score 51; DB 2; Length 1013;
Best Local Similarity 69.2%; Pred. No. 3e+02;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 10 IPEEKKEEKLQ 22
Db 129 LPPEKEEELQ 141

RESULT 15
Q91292
ID Q91292 PRELIMINARY; PRT; 1045 AA.
AC Q91292;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Integrin.
OS Pleurodeles waltlii (Iberian ribbed newt).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae;  
 OC Pleurodeles.  
 OX NCBI\_TaxID=8319;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95377519; PubMed=7649360; DOI=10.1006/dbio.1995.1212;  
 RA Alfandari D., Whittaker C.A., Desimone D.W., Darribere T.;  
 RT "Integrin alpha v subunit is expressed on mesodermal cell surfaces  
 during amphibian gastrulation.";  
 RL Dev. Biol. 170:249-261(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Alfandari D.R.;  
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -1- SIMILARITY: Belongs to the integrin alpha chain family.  
 DR EMBL; X81108; CAA57014.1; -.  
 DR PIR; S60571; S60571.  
 DR HSP; P06756; IL5G.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0008305; C:integrin complex; IEA.  
 DR GO; GO:000515; P:protein binding; IEA.  
 DR GO; GO:0007160; P:cell-matrix adhesion; IEA.  
 DR GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.  
 DR Pfam; PF00357; Integrin alpha; 1.  
 DR PRINTS; PR01185; INTEGRINA.  
 DR SMART; SM00191; Int alpha; 5.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; UNKNOWN\_1.  
 KW Cell adhesion; Integrin; Transmembrane.  
 SQ SEQUENCE 1045 AA; 115421 MW; 4DB34B766B0C648E CRC64;

Query Match 44.7%; Score 51; DB 2; Length 1045;  
 Best Local Similarity 42.9%; Pred.No. 3.1e+02;  
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 1 KLGFFAHKKIPESEKREKLE 21  
 | | | | | : | | | : | | |  
 Db 1014 KFGFFKRVPPQEQREQLQ 1034

Search completed: April 6, 2005, 12:15:39  
 Job time : 5.25607 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:51 ; Search time 2.7823 Seconds  
(without alignments)  
3058.161 Million cell updates/sec

Title: US-09-647-544-7

Perfect score: 114

Sequence: 1 KLGFPAKKIPBEKREKLEQ 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_16Dec04.\*

1: Geneseqp1980s.\*

2: Geneseqp1990s.\*

3: Geneseqp2000s.\*

4: Geneseqp2001s.\*

5: Geneseqp2002s.\*

6: Geneseqp2003as.\*

7: Geneseqp2003bs.\*

8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114	100.0	22	3 AAY32244	Human int
2	114	100.0	1049	5 ADR41424	Human CD-
3	114	100.0	1132	3 AAY32243	Human int
4	114	100.0	1167	3 AAY32242	Human int
5	114	100.0	1167	4 AAB64584	Human sec
6	114	100.0	1167	6 ABP99490	Human sec
7	114	100.0	1167	6 ABR00964	Human gen
8	114	100.0	1167	6 ADA44026	Human sec
9	114	100.0	1167	8 ADQ19290	Human sof
10	52	45.6	344	7 ADC97404	E. faeciu
11	51	44.7	121	4 ABG26508	Novel hum
12	51	44.7	147	4 ABG11269	Novel hum
13	51	44.7	502	4 ABG07742	Novel hum
14	51	44.7	856	4 AAU35717	Helicobac
15	51	44.7	856	4 ABU30742	Protein e
16	50	43.9	32	2 AAR80955	Integrin
17	50	43.9	156	4 ABG18894	Novel hum
18	50	43.9	357	5 ABB48958	Listeria
19	50	43.9	1048	5 AAU76335	Human ant
20	50	43.9	1048	6 ABU03549	Angiogene
21	50	43.9	1048	6 AAO27098	Human int
22	50	43.9	1048	6 ABB82767	Human alp
23	50	43.9	1048	7 ADD48891	Human pro
24	50	43.9	1061	4 ABG18895	Novel hum
25	49.5	43.4	1388	6 ABU38696	Human nuc

26	49.5	43.4	1308	7 ADJ69333	Human hea
27	49	43.0	130	4 AAO00651	Human pol
28	49	43.0	413	5 ADH32817	Yeast sMO
29	49	43.0	486	3 AAG53097	Arabidops
30	49	43.0	584	3 AAG53096	Arabidops
31	49	43.0	615	3 AAG53095	Arabidops
32	49	43.0	751	2 AAW13491	Helicobac
33	49	43.0	856	4 AAU35896	Helicobac
34	48.5	42.5	407	8 ADP49328	S pyrogen
35	48.5	42.5	723	4 ABG28392	Novel hum
36	48	42.1	573	2 AAR82026	Squalene
37	48	42.1	573	7 ADB85174	Rat equal
38	48	42.1	1464	4 ABB61199	Drosophil
39	48	42.1	1526	7 ADE61230	Rat Prote
40	48	42.1	1526	7 ADD44993	Rat Prote
41	47.5	41.7	2343	2 AAW80989	Canine fa
42	47.5	41.7	2343	3 AAY57846	Canine fa
43	47	41.2	136	4 ABG14329	Novel hum
44	47	41.2	177	3 AAG53177	Arabidops
45	47	41.2	177	3 AAG14466	Arabidops

## ALIGNMENTS

### RESULT 1

AAU32244

ID AAY32244 standard; peptide; 22 AA.

XX AC AAY32244;

XX AC AAY32244;

DT 15-FEB-2000 (first entry)

XX Human integrin subunit alpha-10 cytoplasmic domain peptide.

DE Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;

XX Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;

KW osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;

KW therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker.

XX Homo sapiens.

OS Homo sapiens.

XX WO9951639-A1.

PN 14-OCT-1999.

FD 31-MAR-1999; 99WO-SE000544.

XX 02-APR-1998; 98SE-00001164.

XX 28-JAN-1999; 99SE-00000319.

XX (ACTI-) ACTIVE BIOTECH AB.

XX Lundgren-Akerlund E;

XX WPI; 2000-052639/04.

XX New isolated integrin subunit alpha-10, used as a marker or target

PT molecule for cells during development, regeneration and pathological

PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or

PT inflammation.

XX Claim 21; Page 53; 90pp; English.

XX This sequence represents a fragment of novel human chondrocyte integrin

XX subunit alpha-10 (ISa10, see AAY32242), corresponding to the C-terminal

XX cytoplasmic domain of the protein. The invention relates to a recombinant

CC association with subunit beta (especially beta-1). The integrin in

CC heterodimer, or the subunit alpha-10, or a fragment of it such as the

CC present sequence, can be used as a marker or target of all types of

CC cells, e.g. of chondrocytes, osteoblasts and fibroblasts. They can also

CC be used: for treating pathological conditions involving ISa10, such as

CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis; for

CC detecting the formation of cartilage during embryonal development,  
 CC physiological or therapeutic regeneration of cartilage, or detecting  
 CC regeneration of cartilage or chondrocytes during transplantation of  
 CC cartilage or chondrocytes; for selection and analysis or for sorting,  
 CC isolating or purification of chondrocytes and for in vitro studies of  
 CC differentiation of chondrocytes; and as a target for anti-adhesive drugs  
 CC or molecules in tendon, ligament, skeletal muscle or other tissues where  
 CC adhesion impairs the function of the tissue (all claimed). ISa10 binding  
 CC entities can be used to determine the differentiation-state of cells  
 CC during embryonic development, angiogenesis or development of cancer, in  
 CC pathological conditions such as rheumatoid arthritis, osteoarthritis or  
 CC cancer, in tissue regeneration or in therapeutic and physiological  
 CC regeneration of cartilage (claimed). A vaccine comprising the integrin  
 CC heterodimer or subunit alpha-10 is also claimed. ISa10 polynucleotides,  
 CC vectors, host cells and methods of producing recombinant ISa10 are also  
 CC claimed

XX Sequence 22 AA;  
 SQ

Query Match 100.0%; Score 114; DB 3; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-09;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGFPAHKKIPEEKREKLEQ 22  
 |||||  
 Db 1 KLGFPAHKKIPEEKREKLEQ 22  
 |||||

RESULT 2  
 ADR41424  
 ID ADR41424 standard; protein; 1049 AA.  
 XX  
 AC ADR41424;  
 DT 07-OCT-2004 (first entry)  
 XX Human CD-like molecule HA0AD02, SEQ ID NO:223.  
 DE Human; CD-like molecule; cluster of differentiation; diagnosis;  
 XX prevention; immune disorder; immunodeficiency; autoimmune disorder;  
 KW blood-related disorder; haematological disorder; haemostatic disorder;  
 KW thrombolytic disorder; hyperproliferative disorder; cancer; tumour;  
 KW apoptotic disorder; cardiovascular disorder; respiratory disorder;  
 KW angiogenic disorder; neovascularisation; neurological disorder;  
 KW endocrine disorder; reproductive system disorder; infectious disease;  
 KW gastrointestinal disorder; drug screening; tissue regeneration;  
 KW chemotaxis; gene therapy; antibody therapy; drug targeting;  
 KW chromosome mapping; forensic analysis; immunophenotyping; cytostatic;  
 KW haemostatic; tranquiliser; vulnerary; antiinflammatory; nephrotropic;  
 KW cardiant; anti-allergic; anti-HIV; antirheumatic; antiarthritic;  
 KW antipneumatic; immunosuppressive; vasotropic; neuroprotective;  
 KW antithyroid; thyromimetic; gynaecological; virucide; hepatotropic;  
 KW antibacterial; dermatological; chromosome iq21.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200226930-A2.  
 XX  
 PD 04-APR-2002.  
 XX  
 PF 25-SEP-2001; 2001WO-029838.  
 XX  
 PR 26-SEP-2000; 2000US-0235484P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 FI Rosen CA, Birse CE;  
 XX  
 DR WPI; 2002-405050/43.  
 DR N-PSDB; ADR41248.  
 XX  
 XX Novel polynucleotides and polypeptides useful for treating, preventing or  
 PT ameliorating cardiovascular, renal, neurovascular, and autoimmune

PT disorders.  
 XX Claim 11; SEQ ID NO 223; 1243pp; English.  
 PS  
 XX The invention relates to 167 novel human CD (cluster of differentiation)-  
 CC like molecules (ADR41388-ADR41563) and to cDNAs encoding them (seqid11)-  
 CC Sequence 1049 AA;  
 SQ

Query Match 100.0%; Score 114; DB 5; Length 1049;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-07;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGFPAHKKIPEEKREKLEQ 22  
 |||||  
 Db 1028 KLGFPAHKKIPEEKREKLEQ 1049  
 |||||

RESULT 3  
 AAY32243  
 ID AAY32243 standard; protein; 1132 AA.  
 XX  
 AC AAY32243;  
 DT 15-FEB-2000 (first entry)  
 XX Human integrin subunit alpha-10 splice variant.  
 DE Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;  
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;  
 KW therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker;  
 KW splice variant.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..22  
 FT Protein /note= "signal peptide"  
 FT Protein 23..1132  
 FT Protein /note= "mature protein"  
 XX  
 PN WO9951639-A1.  
 XX  
 PD 14-OCT-1999.  
 XX  
 PF 31-MAR-1999; 99WO-SE000544.  
 XX  
 PR 02-APR-1998; 98SE-00001164.  
 PR 28-JAN-1999; 99SE-00000319.  
 XX  
 PA (ACTI-) ACTIVE BIOTECH AB.  
 XX  
 FI Lundgren-Akerlund B;  
 XX  
 DR WPI; 2000-052639/04.  
 DR N-PSDB; AAZ34720.  
 XX  
 XX New isolated integrin subunit alpha-10, used as a marker or target  
 PT molecule for cells during development, regeneration and pathological  
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or  
 PT inflammation.  
 XX  
 PS Claim 1; Page 43-48; 90pp; English.  
 XX

This sequence represents a splice variant of novel human chondrocyte  
 CC integrin subunit alpha-10 (ISa10). It is identical to ISa10 (see  
 CC AAY32242) except for deletion of amino acids 975-986. The invention  
 CC relates to a recombinant or isolated integrin heterodimer comprising the  
 CC alpha10 subunit in association with subunit beta (especially beta-1). The  
 CC heterodimer, subunit alpha-10 or splice variant can be used as a marker  
 CC or target of all types of cells, e.g. of chondrocytes, osteoblasts and  
 CC fibroblasts. They can also be used: for treating pathological conditions  
 CC involving ISa10, such as damage to cartilage, trauma, rheumatoid

CC arthritis or osteoarthritis; for detecting the formation of cartilage  
 CC during embryonal development, physiological or therapeutic repair of  
 CC cartilage, or detecting regeneration of cartilage or chondrocytes during  
 CC transplantation of cartilage or chondrocytes; for selection and analysis  
 CC or for sorting, isolating or purification of chondrocytes and for in  
 CC vitro studies of differentiation of chondrocytes; and as a target for  
 CC anti-adhesive drugs or molecules in tendon, ligament, skeletal muscle or  
 CC other tissues where adhesion impairs the function of the tissue (all  
 CC claimed). ISa10 binding entities can be used to determine the  
 CC differentiation-state of cells during embryonic development, angiogenesis  
 CC or development of cancer, in pathological conditions such as rheumatoid  
 CC arthritis, osteoarthritis or cancer, in tissue regeneration or in  
 CC therapeutic and physiological repair of cartilage (claimed). A  
 CC vaccine comprising the integrin heterodimer or subunit alpha-10 is also  
 CC claimed. ISa10 polynucleotides, vectors, host cells and methods of  
 CC producing recombinant ISa10 are also claimed  
 XX  
 SQ Sequence 1132 AA;

Query Match 100.0%; Score 114; DB 3; Length 1132;  
 Best Local Similarity 100.0%; Pred. NO. 3e-07;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGFFAHKKIPEEKREKLEQ 22  
 |||||  
 Db 1111 KLGFFAHKKIPEEKREKLEQ 1132

RESULT 4  
 ID AAY32242  
 XX AAY32242 standard; protein; 1167 AA.  
 AC AAY32242;  
 XX  
 DT 15-FEB-2000 (first entry)  
 DE Human integrin subunit alpha-10.  
 XX  
 KW Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;  
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation;  
 KW therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker.  
 XX  
 OS Homo sapiens.

XX Key Location/Qualifiers  
 FH Peptide 1..22  
 FT /note= "signal peptide"  
 FT Protein 23..1145  
 FT /note= "mature protein"  
 FT Domain 23..1120  
 FT /note= "extracellular domain"  
 FT Modified-site 98  
 FT /note= "N-glycosylated"  
 FT Domain 162..359  
 FT /note= "I-domain"  
 FT Modified-site 336  
 FT /note= "N-glycosylated"  
 FT Modified-site 364  
 FT /note= "N-glycosylated"  
 FT Binding-site 494..502  
 FT /note= "cation binding site motif"  
 FT Binding-site 558..566  
 FT /note= "cation binding site motif"  
 FT Binding-site 620..628  
 FT /note= "cation binding site motif"  
 FT Modified-site 733  
 FT /note= "N-glycosylated"  
 FT Modified-site 839  
 FT /note= "N-glycosylated"  
 FT Modified-site 921  
 FT /note= "N-glycosylated"  
 FT Modified-site 1018  
 FT /note= "N-glycosylated"

FT Modified-site 1039  
 FT /note= "N-glycosylated"  
 FT Domain 1121..1145  
 FT /note= "transmembrane domain"  
 FT Domain 1122..1167  
 FT /note= "cytoplasmic domain, specifically claimed in Claim 21"  
 XX  
 PN WO9951639-A1.  
 XX  
 PD 14-OCT-1999.  
 XX  
 PF 31-MAR-1999; 99WO-SB0000544.  
 XX  
 PR 02-APR-1998; 98SE-00001164.  
 PR 28-JAN-1999; 99SE-00000319.  
 XX  
 PA (ACTI-) ACTIVE BIOTECH AB.  
 XX  
 PI Lundgren-Akerlund E;  
 XX  
 DR WPI; 2000-052639/04.  
 DR N-PSDB; AAZ34719.  
 XX  
 PT New isolated integrin subunit alpha-10, used as a marker or target  
 PT molecule for cells during development, regeneration and pathological  
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or  
 PT inflammation.  
 XX  
 PS Claim 1; Fig 6; 90pp; English.

CC This sequence represents novel human chondrocyte integrin subunit alpha-10 (ISa10). A splice variant is given in AAY32243. The invention relates to a recombinant or isolated integrin heterodimer comprising the alpha10 subunit in association with subunit beta (especially beta-1). The heterodimer and the subunit alpha-10 can be used as markers or targets of all types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts. They can also be used for treating pathological conditions involving ISa10, such as damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis; for detecting the formation of cartilage during embryonal development, physiological or therapeutic repair of cartilage, or detecting regeneration of cartilage or chondrocytes during transplantation of cartilage or chondrocytes; for selection and analysis or for sorting, isolating or purification of chondrocytes and for in vitro studies of differentiation of chondrocytes; and as a target for anti-adhesive drugs or molecules in tendon, ligament, skeletal muscle or other tissues where adhesion impairs the function of the tissue (all claimed). ISa10 binding entities can be used to determine the differentiation-state of cells during embryonic development, angiogenesis or development of cancer, in pathological conditions such as rheumatoid arthritis, osteoarthritis or cancer, in tissue regeneration or in therapeutic and physiological repair of cartilage (claimed). A vaccine comprising the integrin heterodimer or subunit alpha-10 is also claimed. ISa10 polynucleotides, vectors, host cells and methods of producing recombinant ISa10 are also claimed

XX SQ Sequence 1167 AA;

Query Match 100.0%; Score 114; DB 3; Length 1167;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-07;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGFFAHKKIPEEKREKLEQ 22  
 |||||  
 Db 1146 KLGFFAHKKIPEEKREKLEQ 1167

RESULT 5  
 AAB64584  
 ID AAB64584 standard; protein; 1167 AA.  
 XX  
 AC AAB64584;  
 XX

DT 22-MAR-2001 (first entry)  
XX Human secreted protein #37.  
KW Cytostatic; immunosuppressive; neutropic; neuroprotective; antiviral;  
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KW vulnarary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein.  
XX Homo sapiens.  
XX WO200077197-A1.  
XX 21-DEC-2000.  
XX 01-JUN-2000; 2000WO-US014934.  
XX 11-JUN-1999; 99US-0138599P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX (ROSE/) ROSEN C A.  
XX Rosen CA, Ruben SM, Komatsoulis GA;  
XX WPI; 2001-032312/04.  
XX N-PSDB; AAF32793.  
XX Isolated nucleic acid molecule encoding a human secreted protein is used  
XX in preventing, treating or ameliorating a medical condition.  
XX Claim 11; Page 496-500; 558pp; English.  
XX Sequences AAB64549-B64594 represent the amino acid sequences of 47 human  
XX secreted proteins encoded by the genes AAF32757-F32803. The genes and  
XX proteins are useful for preventing, ameliorating or treating medical  
XX conditions, e.g. by protein or gene therapy. The genes are isolated from  
XX a range of human tissues disclosed in the specification. The nucleic  
XX acids, proteins, antibodies and (ant)agonists are useful in the  
XX diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
XX ovarian cancer, and other cancers of the adrenal gland, bone, bone  
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b)  
XX immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
XX anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
XX cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
XX; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
XX infectious diseases such as viral, bacterial, fungal and parasitic  
XX infections  
XX Sequence 1167 AA;  
XX  
XX Query Match 100.0%; Score 114; DB 4; Length 1167;  
XX Best Local Similarity 100.0%; Pred. No. 3.1e-07;  
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1 KLGFPAHKKIPPEEKREKLEQ 22  
XX 1146 KLGFPAHKKIPPEEKREKLEQ 1167  
XX  
XX RESULT 6  
XX ABP99490  
XX ID ABP99490 standard; protein; 1167 AA.  
XX AC ABP99490;  
XX 26-MAR-2003 (first entry)  
XX Human secreted protein SEQ ID NO 434.  
XX Human; secreted protein; neutropic; neuroprotective; cytostatic;  
KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;  
KW

KW vulnarary; antibacterial; antiparkinsonian; antisickling; antianemic;  
KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;  
KW antiinflammatory; antiallergic; antidiabetic; antiulcer; anticonvulsant;  
KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;  
KW cardiovascular disorder; neurological disease; nephrotropic;  
XX gene therapy.  
XX Homo sapiens.  
XX WO200277186-A2.  
XX 03-OCT-2002.  
XX 26-MAR-2002; 2002WO-US009188.  
XX 27-MAR-2001; 2001US-0278650P.  
XX 12-SEP-2001; 2001US-00950082.  
XX 12-SEP-2001; 2001US-00950083.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Ruben SM;  
XX WPI; 2003-040583/03.  
XX N-PSDB; ABZ66911.  
XX New human secreted proteins encoded by genes contained in cDNA clones  
XX (e.g. HGACAC19), useful for preventing, treating or diagnosing e.g. AIDS,  
XX multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or  
XX West Nile fever.  
XX Claim 1; Page 1395-1398; 2423pp; English.  
XX The invention relates to novel human genes (ABZ66891-ABZ68209) and the  
XX encoded secreted proteins (ABP99470-ABP99872) useful for preventing,  
XX treating or ameliorating medical conditions e.g. by protein or gene  
XX therapy. The genes are isolated from a range of human tissues disclosed  
XX in the specification. The nucleic acids, proteins, antibodies and  
XX (ant)agonists are useful in the diagnosis, treatment and prevention of:  
XX (a) cancer, e.g. breast and ovarian cancer and other cancers of the  
XX adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
XX lung or urogenital; (b) immune disorders e.g. Addison's disease,  
XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
XX arthritis and ulcerative colitis; (c) cardiovascular disorders such as  
XX myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.  
XX cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,  
XX bacterial, fungal and parasitic infections  
XX Sequence 1167 AA;  
XX  
XX Query Match 100.0%; Score 114; DB 6; Length 1167;  
XX Best Local Similarity 100.0%; Pred. No. 3.1e-07;  
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1 KLGFPAHKKIPPEEKREKLEQ 22  
XX 1146 KLGFPAHKKIPPEEKREKLEQ 1167  
XX  
XX RESULT 7  
XX ABR00964  
XX ID ABR00964 standard; protein; 1167 AA.  
XX AC ABR00964;  
XX 12-MAY-2003 (first entry)  
XX Human gene 18-encoded secreted protein HA0AG15, SEQ ID NO:445.  
XX Human; secreted protein; cancer; tumour; hyperproliferative disorder;  
KW autoimmune disorder; inflammation; angioecic diseases; AIDS;  
KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;  
KW

KW drug screening; chromosome identification; chromosome mapping;  
 KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;  
 KW antianemic; vulnerary; chromosome iq21.  
 OS Homo sapiens.  
 XX WO200277013-A2.  
 XX 03-OCT-2002.  
 XX 26-MAR-2002; 2002WO-US009370.  
 XX 27-MAR-2001; 2001US-0278650P.  
 PR 12-SEP-2001; 2001US-00950082.  
 PR 12-SEP-2001; 2001US-00950083.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Ruben SM;  
 PI WPI; 2003-040578/03.  
 DR N-PSDB; ABZ73298.  
 DR New human secreted proteins and nucleic acids, useful for detecting or  
 PT treating cancer or other hyperproliferative disorders, autoimmune  
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.  
 XX Claim 13; Page 1387-1390; 2474pp; English.  
 PS ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted  
 CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.  
 CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The  
 CC invention also encompasses antibodies specific for the secreted proteins,  
 CC the use of the secreted proteins in drug screening and recombinant  
 CC vectors and host cells comprising a nucleic acid of the invention. The  
 CC secreted proteins are thought to be involved in biological activities  
 CC associated with cellular signalling, cellular differentiation, cell  
 CC migration, prohormone activation and neurotransmitter activity. The  
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody  
 CC fragments specific for the secreted proteins, and modulators of protein  
 CC activity are useful for the diagnosing or treating cancers or other  
 CC hyperproliferative disorders. Additionally, the secreted proteins and  
 CC their nucleic acids may also be used in the treatment of autoimmune  
 CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS  
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote  
 CC wound healing. Nucleic acids of the invention may be used for chromosome  
 CC identification, chromosome mapping, in gene therapy, for identifying  
 CC individuals from minute biological samples, as hybridisation probes, and  
 CC as molecular weight markers. The present sequence represents a human  
 CC secreted protein of the invention  
 XX SQ Sequence 1167 AA;  
 Query Match 100.0%; Score 114; DB 6; Length 1167;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-07;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 KLGFFAHKKIPEEKREKLEQ 22  
 |||||  
 Db 1146 KLGFFAHKKIPEEKREKLEQ 1167  
 RESULT 8  
 ADA44026  
 ID ADA44026 standard; protein; 1167 AA.  
 XX ADA44026;  
 XX 20-NOV-2003 (first entry)  
 DT Human secreted protein SEQ ID 214.  
 DE Gene therapy; human; Antidiabetic; Anorectic; Ophthalmological;  
 KW

KW Neuroprotective; Cerebroprotective; Antianemic.  
 XX Homo sapiens.  
 OS WO2003000865-A2.  
 XX 03-JAN-2003.  
 XX 26-MAR-2002; 2002WO-US009105.  
 XX 27-MAR-2001; 2001US-0278650P.  
 PR 12-SEP-2001; 2001US-00950082.  
 PR 12-SEP-2001; 2001US-00950083.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Ruben SM;  
 PI WPI; 2003-184045/18.  
 DR N-PSDB; ADA43832.  
 DR A human secreted protein and nucleic acids useful for preparing a  
 PT diagnostic or pharmaceutical composition for diagnosing or treating  
 PT diabetes or conditions related to diabetes, e.g. hyperglycemia, obesity,  
 PT retinopathy, neuropathy.  
 XX Claim 1; SEQ ID NO 214; 701pp; English.  
 PS The invention relates to novel genes and their fragments which are useful  
 CC for preventing, treating or ameliorating medical conditions e.g. by  
 CC protein or gene therapy. The genes are isolated from a range of human  
 CC tissues disclosed in the specification. The nucleic acids and proteins  
 CC are useful in the diagnosis, treatment and prevention of conditions  
 CC related to diabetes, e.g. hyperglycaemia, obesity, retinopathy,  
 CC polynuropathy, atherosclerosis, anaemia, stroke, gangrene, impotence,  
 CC infection, cataract, renal disorders, or endocrine disorders. The present  
 CC sequence was used to illustrate the invention.  
 XX SQ Sequence 1167 AA;  
 Query Match 100.0%; Score 114; DB 6; Length 1167;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-07;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 KLGFFAHKKIPEEKREKLEQ 22  
 |||||  
 Db 1146 KLGFFAHKKIPEEKREKLEQ 1167  
 RESULT 9  
 ADQ19290  
 ID ADQ19290 standard; protein; 1167 AA.  
 XX ADQ19290;  
 XX 26-AUG-2004 (first entry)  
 DT Human soft tissue sarcoma-upregulated protein - SEQ ID 2109.  
 DE soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.  
 KW Homo sapiens.  
 OS WO2004048938-A2.  
 XX 10-JUN-2004.  
 XX 26-NOV-2003; 2003WO-US038193.  
 PF 26-NOV-2002; 2002US-0429739P.  
 XX (PROT-) PROTEIN DESIGN LABS INC.  
 XX

```

PI Aziz N, Gineburg WM, Zlotnik A;
XX WPI; 2004-441208/41.
XX
XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
XX Example 2; SEQ ID NO 2109; 210pp; English.
XX
XX The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC protein of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
XX Sequence 1167 AA;

Query Match 100.0%; Score 114; DB 8; Length 1167;
Best Local Similarity 100.0%; Pred. NO. 3.1e-07;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGFFFAHKKIPPEEKREKLEQ 22
DB 1146 KLGFFFAHKKIPPEEKREKLEQ 1167

RESULT 10
ADC97404
ID ADC97404 standard; protein; 344 AA.
XX
AC ADC97404;
XX
XX 01-JAN-2004 (first entry)
XX
XX E. faecium protein sequence SEQ ID 7031.
XX
XX Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
XX abdominal-pelvic infection.
XX
XX Enterococcus faecium.
XX
XX US6583275-B1.
XX
XX 24-JUN-2003.
XX
XX 30-JUN-1998; 98US-00107532.
XX
XX 02-JUL-1997; 97US-0051571P.
XX
XX 14-MAY-1998; 98US-0085598P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI; 2003-799836/75.
XX
XX N-PSDB; ADC93750.
XX
XX New isolated nucleic acid derived from Enterococcus faecium encoding an
PT Enterococcus faecium polypeptide useful for detection, prevention and
PT treatment of a pathological condition resulting from a bacterial
PT infection.
XX
XX Example 1; SEQ ID NO 7031; 243pp; English.
PS

```

---

```

XX The invention relates to an isolated nucleic acid derived from
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
CC one of 10 fully defined sequences given in the (or comprising 40
CC sequential nucleotides chosen from any of the nucleic acids, its
CC complement or sequences hybridising to it). Also included are a
CC recombinant vector comprising the nucleic acid operably linked to
CC transcription regulatory element, a cell comprising the vector and a
CC single-stranded probe comprising the nucleic acid. The nucleic acids are
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
CC The nucleic acids are useful for diagnosing pathological conditions
CC resulting from E. faecium bacterial infection (e.g. urinary tract
CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
CC infection) and for screening drugs such as agonists and antagonists. The
CC nucleic acid is useful for recombinant production of Candida albicans -
CC derived peptides or antisense polypeptides. Pharmaceutical compositions
CC and vaccines containing the nucleic acid are useful for preventing or
CC treating Enterococcus faecium infections. The present sequence represents
CC one if the disclosed E. faecium proteins.
XX
XX Sequence 344 AA;

Query Match 45.6%; Score 52; DB 7; Length 344;
Best Local Similarity 47.1%; Pred. No. 47;
Matches 8; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 KLGFFFAHKKIPPEEKRE 17
DB 13 RMFFHYKKIPPEEREQ 29

RESULT 11
ABG26508
ID ABG26508 standard; protein; 121 AA.
XX
AC ABG26508;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #26499.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSB-) HYSBQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS90695.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 56867; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC

```



CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 121 AA;

Query Match 44.7%; Score 51; DB 4; Length 121;  
 Best Local Similarity 55.6%; Pred. No. 23;  
 Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 Qy 5 FAHKKIPPEEKREKLEQ 22  
 | : ||| ||| : || : |  
 Db 5 FLYKKIKEEEEEEEEE 22

RESULT 12  
 ABG11269  
 ID ABG11269 standard; protein; 147 AA.

XX AC ABG11269;  
 XX DT 18-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #11260.  
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX FN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US008631.  
 XX PR 31-MAR-2000; 2000US-00540217.  
 XX PR 23-AUG-2000; 2000US-00649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Drmanac RT, Liu C, Tang YT;  
 XX DR WPI; 2001-639362/73.  
 XX DR N-PSDB; AAS75456.  
 XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX PS Claim 20; SEQ ID NO 41628; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a

CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 147 AA;

Query Match 44.7%; Score 51; DB 4; Length 147;  
 Best Local Similarity 55.6%; Pred. No. 28;  
 Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 Qy 5 FAHKKIPPEEKREKLEQ 22  
 | : ||| ||| : || : |  
 Db 5 FLYKKIKEEEEEEEEE 22

RESULT 13  
 ABG07742  
 ID ABG07742 standard; protein; 502 AA.

XX AC ABG07742;  
 XX DT 13-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #7733.  
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX FN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US008631.  
 XX PR 31-MAR-2000; 2000US-00540217.  
 XX PR 23-AUG-2000; 2000US-00649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Drmanac RT, Liu C, Tang YT;  
 XX DR WPI; 2001-639362/73.  
 XX DR N-PSDB; AAS71929.  
 XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX PS Claim 20; SEQ ID NO 38101; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a

CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 XX SQ Sequence 502 AA;

Query Match 44.7%; Score 51; DB 4; Length 502;  
 Best Local Similarity 55.6%; Pred. No. 95;  
 Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 FAHKKIPPEEKREKLEQ 22  
 :|||||:|||||:  
 Db 5 FLYKKKEEEEEEE 22

RESULT 14  
 AAU35717  
 ID AAU35717 standard; protein; 856 AA.

XX AC AAU35717;

XX DT 14-FEB-2002 (first entry)

XX DE Helicobacter pylori cellular proliferation protein #30.

XX KW Antisense; prokaryotic cellular proliferation protein; antibiotic;  
 XX KB antibacterial; drug design.

XX OS Helicobacter pylori.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US009180.

XX PR 21-MAR-2000; 2000US-0191078P.

XX PR 23-MAY-2000; 2000US-0206848P.

XX PR 26-MAY-2000; 2000US-0207727P.

XX PR 23-OCT-2000; 2000US-0242578P.

XX PR 27-NOV-2000; 2000US-0253625P.

XX PR 22-DEC-2000; 2000US-0257931P.

XX PR 16-FEB-2001; 2001US-0269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

XX DR N-PSDB; AAS53576.

XX PT New polynucleotides for the identification and development of

XX PT antibiotics, comprise sequences of antisense nucleic acids.

XX Example 3; SEQ ID NO 11310; 511bp; English.

XX The invention relates to antisense inhibitors of genes essential to

XX prokaryotic cellular proliferation, their use in identifying the genes,

CC useful for the identification of potential new targets for antibiotic  
 CC development. The antisense nucleic acids can also be used to identify  
 CC proteins used in proliferation, to express these proteins, and to obtain  
 CC antibodies capable of binding to the expressed proteins. The proteins can  
 CC be used to screen compounds in rational drug discovery programmes. The  
 CC antisense nucleic acid sequence is also useful to screen for homologous  
 CC nucleic acids which are required for cell proliferation in a wide variety  
 CC of organisms. The present sequence represents an essential prokaryotic  
 CC cellular proliferation protein. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 XX SQ Sequence 856 AA;

Query Match 44.7%; Score 51; DB 4; Length 856;  
 Best Local Similarity 45.0%; Pred. No. 1.6e+02;  
 Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 3 GFFAHHKIPPEEKREKLEQ 22  
 :|||||:|||||:  
 Db 496 GREYSKIPENKKKEELQR 515

RESULT 15

ABU30742

ID ABU30742 standard; protein; 856 AA.

XX AC ABU30742;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #16269.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Helicobacter pylori.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX DR N-PSDB; ACA34612.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 58666; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

XX the 6213 antisense sequences given in the specification where expression

XX of the nucleic acid inhibits proliferation of a cell. Also included are:

XX (1) a vector comprising a promoter operably linked to the nucleic acid

XX encoding a polypeptide whose expression is inhibited by the antisense

XX nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX polypeptide or its fragment whose expression is inhibited by the

XX antisense nucleic acid; (4) an antibody capable of specifically binding

XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2005, 12:17:21 ; Search time 2.74785 Seconds  
(without alignments)  
2658.060 Million cell updates/sec

Title: US-09-647-544-7

Perfect score: 114

Sequence: 1 KLGFFAHKKIPPEEKREKLEQ 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1418010 seqs, 331997259 residues

Total number of hits satisfying chosen parameters: 1418010

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114	100.0	1167	US-10-741-601-531	Sequence 531, App
2	114	100.0	1177	US-10-741-601-532	Sequence 532, App
3	52.5	46.1	284	US-10-437-963-148869	Sequence 148869, App
4	52.5	46.1	1170	US-10-437-963-189133	Sequence 189133, App
5	52	45.6	137	US-10-424-599-277893	Sequence 277893, App
6	52	45.6	577	US-10-424-599-277888	Sequence 277888, App
7	51	44.7	856	US-09-812-350-2	Sequence 2, Appli
8	51	44.7	856	US-09-815-242-11310	Sequence 11310, A
9	51	44.7	856	US-10-282-122A-58666	Sequence 58666, A
10	50	43.9	1048	US-09-920-267C-9	Sequence 9, Appli
11	50	43.9	1048	US-10-211-462-189	Sequence 189, App
12	50	43.9	1048	US-10-720-323-9	Sequence 9, Appli
13	50	43.9	1048	US-10-482-029-20	Sequence 20, Appli

14	49.5	43.4	1388	16	US-10-408-765A-1139	Sequence 1139, Ap
15	49	43.0	107	16	US-10-437-963-204563	Sequence 204563,
16	49	43.0	108	15	US-10-424-599-155879	Sequence 155879, Ap
17	49	43.0	413	14	US-10-083-357-1275	Sequence 1275, Ap
18	49	43.0	841	16	US-10-437-963-187432	Sequence 187432,
19	49	43.0	856	9	US-09-815-242-11489	Sequence 11489, A
20	49	43.0	856	15	US-10-335-977-7660	Sequence 7660, Ap
21	48	42.1	118	16	US-10-437-963-165708	Sequence 165708,
22	48	42.1	288	15	US-10-424-599-284529	Sequence 284529,
23	48	42.1	565	15	US-10-424-599-275593	Sequence 275593,
24	48	42.1	573	14	US-10-205-194-55	Sequence 55, Appl
25	47.5	41.7	126	16	US-10-437-963-115949	Sequence 115949,
26	47	41.2	49	15	US-10-424-599-184147	Sequence 184147,
27	47	41.2	122	15	US-10-335-977-8097	Sequence 8097, Ap
28	47	41.2	572	15	US-10-425-114-57970	Sequence 57970, A
29	47	41.2	646	15	US-10-425-114-70985	Sequence 70985, A
30	47	41.2	860	15	US-10-389-566-1868	Sequence 1868, Ap
31	47	41.2	885	16	US-10-437-963-160466	Sequence 160466,
32	46	40.4	118	16	US-10-767-701-50727	Sequence 50727, A
33	46	40.4	195	15	US-10-424-599-175252	Sequence 175252,
34	46	40.4	221	16	US-10-437-963-175346	Sequence 175346,
35	46	40.4	274	15	US-10-424-599-159588	Sequence 159588,
36	46	40.4	308	15	US-10-282-122A-73679	Sequence 73679, A
37	46	40.4	308	17	US-10-472-928-3922	Sequence 3922, Ap
38	46	40.4	363	16	US-10-437-963-182343	Sequence 182343,
39	46	40.4	479	9	US-09-971-309-64	Sequence 64, Appl
40	46	40.4	479	16	US-10-828-924-63	Sequence 63, Appl
41	46	40.4	753	14	US-10-032-585-7608	Sequence 7608, Ap
42	46	40.4	764	15	US-10-425-114-54331	Sequence 54331, A
43	46	40.4	1499	16	US-10-437-963-112408	Sequence 112408,
44	46	40.4	1603	16	US-10-828-924-62	Sequence 62, Appl
45	45	39.5	109	15	US-10-001-192A-31	Sequence 31, Appl

#### ALIGNMENTS

RESULT 1  
US-10-741-601-531  
; Sequence 531, Application US/10741601  
; Publication No. US20040166519A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH STENOSIS, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01500  
; CURRENT APPLICATION NUMBER: US/10/741,601  
; CURRENT FILING DATE: 2003-12-22  
; NUMBER OF SEQ ID NOS: 26415  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 531  
; LENGTH: 1167  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-741-601-531

Query Match 100.0%; Score 114; DB 16; Length 1167;  
Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGFFAHKKIPPEEKREKLEQ 22  
Db 1146 KLGFFAHKKIPPEEKREKLEQ 1167

RESULT 2  
US-10-741-601-532  
; Sequence 532, Application US/10741601  
; Publication No. US20040166519A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH STENOSIS, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001500  
; CURRENT APPLICATION NUMBER: US/10/741.601  
; CURRENT FILING DATE: 2003-12-22  
; NUMBER OF SEQ ID NOS: 26415  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 532  
; LENGTH: 1177  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-741-601-532

Query Match 100.0%; Score 114; DB 16; Length 1177;  
Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGFFAHKKIPEEKREKLEQ 22  
DB 1146 KLGFFAHKKIPEEKREKLEQ 1167

## RESULT 3

US-10-437-963-148869  
; Sequence 148869, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437.963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 148869  
; LENGTH: 284  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_49258C.1.pep  
US-10-437-963-148869

Query Match 46.1%; Score 52.5; DB 16; Length 284;  
Best Local Similarity 57.9%; Pred. No. 45;  
Matches 11; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 2 LGFFAHKKIPEEKREKLEQ 20  
DB 131 LGLF-DRRIPEEEEEKEM 148

## RESULT 4

US-10-437-963-189133  
; Sequence 189133, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437.963

; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 189133  
; LENGTH: 1170  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_85670C.1.pep  
US-10-437-963-189133

Query Match 46.1%; Score 52.5; DB 16; Length 1170;  
Best Local Similarity 57.9%; Pred. No. 1.9e+02;  
Matches 11; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 2 LGFFAHKKIPEEKREKLEQ 20  
DB 1017 LGLF-DRRIPEEEEEKEM 1034

## RESULT 5

US-10-424-599-277893  
; Sequence 277893, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424.599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 277893  
; LENGTH: 137  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_92960C.1.pep  
US-10-424-599-277893

Query Match 45.6%; Score 52; DB 15; Length 137;  
Best Local Similarity 41.2%; Pred. No. 25;  
Matches 7; Conservative 8; Mismatches 2; Indels 0; Gaps 0;

QY 6 AHKKIPEEKREKLEQ 22  
DB 93 SHRRVPEKSEEELEE 109

## RESULT 6

US-10-424-599-277888  
; Sequence 277888, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424.599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 277888  
; LENGTH: 577  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(577)

```
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_92956C.1.pep
US-10-424-599-277888

Query Match      45.6%; Score 52; DB 15; Length 577;
Best Local Similarity 41.2%; Pred. No. 1.le+02;
Matches 7; Conservative 8; Mismatches 2; Indels 0; Gaps 0;

Qy 6 AHKKIPEEKREKLEQ 22
:|::|||:|::|||:
Db 93 SHRVPEKSBEERLEE 109

RESULT 7
US-09-812-350-2
; Sequence 2, Application US/09812350
; Publication No. US20020053097A1
; GENERAL INFORMATION:
; APPLICANT: Lindquist, Susan
; APPLICANT: Queitsch, Christine
; APPLICANT: Vierling, Elizabeth
; TITLE OF INVENTION: Transgenic Plants Containing Heat Shock Protein
; FILE REFERENCE: P01979US2
; CURRENT APPLICATION NUMBER: US/09/812,350
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/190,769
; PRIOR FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: US 60/198,116
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-812-350-2

Query Match      44.7%; Score 51; DB 9; Length 856;
Best Local Similarity 45.0%; Pred. No. 2.le+02;
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 3 GFFAHKKIPEEKREKLEQ 22
|:|::|||:|::|||:
Db 496 GEIYSKIPENKKKEELQR 515

RESULT 8
US-09-815-242-11310
; Sequence 11310, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58666
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-282-122A-58666
```

```
; OTHER INFORMATION:
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11310
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11310

Query Match      44.7%; Score 51; DB 9; Length 856;
Best Local Similarity 45.0%; Pred. No. 2.le+02;
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 3 GFFAHKKIPEEKREKLEQ 22
|:|::|||:|::|||:
Db 496 GEIYSKIPENKKKEELQR 515

RESULT 9
US-10-282-122A-58666
; Sequence 58666, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58666
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-10-282-122A-58666
```





; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Fahy, Eoin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Warlock, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME  
; FILE REFERENCE: 66088.465  
; CURRENT APPLICATION NUMBER: US/10/408,765A  
; CURRENT FILING DATE: 2003-04-04  
; NUMBER OF SEQ ID NOS: 3077  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1139  
; LENGTH: 1388  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-408-765A-1139

Query Match 43.4%; Score 49.5; DB 16; Length 1388;  
Best Local Similarity 57.9%; Pred. No. 5.5e+02;  
Matches 11; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

Qy 4 FFAH---KKIPEEKREEK 19  
Db 366 FFAHLQKVLAEERKQK 384

## RESULT 15

US-10-437-963-204563  
; Sequence 204563, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 204563  
; LENGTH: 107  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_99638C.1.pep  
US-10-437-963-204563

Query Match 43.0%; Score 49; DB 16; Length 107;  
Best Local Similarity 47.4%; Pred. No. 49;  
Matches 9; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KLGFFAHKKIPEEKREEK 19  
Db 82 KLGFSPKPKKKKKKKKK 100

Search completed: April 6, 2005, 12:53:10  
Job time : 3.74785 secs

Page Blank (uspio)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: April 6, 2005, 12:08:52 ; Search time 0.938919 Seconds  
(without alignments)  
1749.117 Million cell updates/sec

Title: US-09-647-544-7

Perfect score: 114

Sequence: 1 KLGFFAHKTIPEEKREKLEQ 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*\*

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*\*
- 5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	45.6	344	4	US-09-107-532A-7031
2	50	43.9	32	1	Sequence 7031, Ap
3	50	43.9	32	5	Sequence 2, Appli
4	50	43.9	1048	4	Sequence 2, Appli
5	50	43.9	1149	4	Sequence 6192, Ap
6	48	42.1	573	2	Sequence 7401, Ap
7	48	42.1	573	3	Sequence 4, Appli
8	47.5	41.7	2343	3	Sequence 7, Appli
9	46	40.4	308	4	Sequence 2, Appli
10	46	40.4	331	4	Sequence 2953, Ap
11	46	40.4	450	4	Sequence 2623, Ap
12	46	40.4	479	3	Sequence 493, Ap
13	46	40.4	479	3	Sequence 64, Appl
14	46	40.4	755	4	Sequence 64, Appl
15	45	39.5	109	3	Sequence 14733 A
16	45	39.5	476	4	Sequence 31, Appl
17	44	38.6	46	1	Sequence 15008, A
18	44	38.6	93	1	Sequence 3, Appli
19	44	38.6	93	1	Sequence 7, Appli
20	44	38.6	93	1	Sequence 16, Appl
21	44	38.6	93	1	Sequence 1, Appli
22	44	38.6	93	4	Sequence 2, Appli
23	44	38.6	200	4	Sequence 3, Appli
24	44	38.6	200	4	Sequence 789, App
25	44	38.6	200	4	Sequence 789, App
26	44	38.6	200	4	Sequence 789, App
27	44	38.6	200	4	Sequence 789, App

28	44	38.6	200	4	US-09-658-824-789	Sequence 789, App
29	44	38.6	207	4	US-09-702-705-1667	Sequence 1667, Ap
30	44	38.6	207	4	US-09-736-457-1667	Sequence 1667, Ap
31	44	38.6	207	4	US-09-614-124B-1667	Sequence 1667, Ap
32	44	38.6	207	4	US-09-671-325-1667	Sequence 1667, Ap
33	44	38.6	207	4	US-09-658-824-1667	Sequence 1667, Ap
34	44	38.6	228	4	US-09-949-016-10496	Sequence 10496, A
35	44	38.6	481	4	US-09-248-796A-17118	Sequence 17118, A
36	44	38.6	745	4	US-09-248-796A-17272	Sequence 17272, A
37	44	38.6	995	4	US-09-362-842-2	Sequence 2, Appli
38	43.5	38.2	765	1	US-08-425-061-19	Sequence 19, Appl
39	43.5	38.2	765	2	US-08-825-886-19	Sequence 19, Appl
40	43.5	38.2	765	4	US-08-989-890-19	Sequence 19, Appl
41	43.5	38.2	900	1	US-08-425-061-20	Sequence 20, Appl
42	43.5	38.2	900	2	US-08-825-886-20	Sequence 20, Appl
43	43.5	38.2	900	4	US-08-989-890-20	Sequence 20, Appl
44	43.5	38.2	914	1	US-08-425-061-21	Sequence 21, Appl
45	43.5	38.2	914	2	US-08-825-886-21	Sequence 21, Appl

## ALIGNMENTS

### RESULT 1

US-09-107-532A-7031

; Sequence 7031, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660

; COMPUTER: PC

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107.532A

; FILING DATE: 30-Jun-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/085,598

; FILING DATE: 14 May 1998

; APPLICATION NUMBER: 60/051571

; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Ariniello, Pamela Deneke

; REGISTRATION NUMBER: 40,489

; REFERENCE/DOCKET NUMBER: GTC-012

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781)893-5007

; TELEFAX: (781)893-8277

; INFORMATION FOR SEQ ID NO: 7031:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 344 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHEICAL: YES

; ORIGINAL SOURCE:

; ORGANISM: Enterococcus faecium

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (B) LOCATION 1...344

; SEQUENCE DESCRIPTION: SEQ ID NO: 7031:

US-09-107-532A-7031



Qy 1 KLGFPAHKKIPPEEKREKLE 21  
:|||||:|:|:|:|:  
Db 1017 RMGFFKVRPPQEQEREQLQ 1037

RESULT 5  
US-09-949-016-7401  
; Sequence 7401, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR FILING DATE: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 7401  
; LENGTH: 1149  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-7401

Query Match 43.9%; Score 50; DB 4; Length 1149;  
Best Local Similarity 38.1%; Pred. No. 84;  
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 1 KLGFPAHKKIPPEEKREKLE 21  
:|||||:|:|:|:|:  
Db 1118 RMGFFKVRPPQEQEREQLQ 1138

RESULT 6  
US-08-745-934-4  
; Sequence 4, Application US/08745934  
; Patent No. 5861496  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Hawkins, Phillip R.  
; TITLE OF INVENTION: HUMAN SQUALENE EPOXIDASE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/745,934  
; FILING DATE: Herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0151 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 855-0555

TELEFAX: (415) 845-4166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 573 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 1083804  
US-08-745-934-4

Query Match 42.1%; Score 48; DB 2; Length 573;  
Best Local Similarity 52.9%; Pred. No. 80;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 LGFFAHKKIPPEEKREE 18  
:|||||:|:|:|:|:  
Db 75 IGFVAKSPPESEKKEQ 91

RESULT 7  
US-09-147-009-7  
; Sequence 7, Application US/09147009  
; Patent No. 6153815  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Process for Raising Squalene Levels in Plants  
; AND DNA SEQUENCES USED THEREFOR  
; NUMBER OF SEQUENCES: 11  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/147,009  
; FILING DATE:  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 573 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Rattus norvegicus  
; TISSUE TYPE: kidney  
; CELL LINE: NRK  
; IMMEDIATE SOURCE:  
; LIBRARY: pCD2 library of H. Okayama  
; CLONE: Tb-1  
; PUBLICATION INFORMATION:  
; AUTHORS: Sakakibara, J.  
; AUTHORS: Watanabe, R.  
; AUTHORS: Kanai, R.  
; AUTHORS: Ono, T.  
; TITLE: Molecular cloning and expression of rat  
; TITLE: squalene epoxidase  
; JOURNAL: J. Biol. Chem.  
; VOLUME: 270  
; ISSUE: 1  
; PAGES: 17-20  
; DATE: 1995  
US-09-147-009-7

Query Match 42.1%; Score 48; DB 3; Length 573;  
Best Local Similarity 52.9%; Pred. No. 80;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 LGFFAHKKIPBEKREE 18  
:|||||:|||||:  
Db 75 IGFFWAKSPPESEKKEQ 91

RESULT 8  
US-09-324-867-2  
; Sequence 2, Application US/09324867A  
; Patent No. 6251632  
; GENERAL INFORMATION:  
; APPLICANT: Lillicrap, David  
; APPLICANT: Cameron, Cherie  
; APPLICANT: No. 6251632ley, Colleen  
; APPLICANT: Horrocks, L. Suzanne Hoyle  
; APPLICANT: Hough, Christine  
; TITLE OF INVENTION: Canine Factor VIII Gene, Protein and Methods of Use  
; FILE REFERENCE: 1669-0010002/JAG/BJD  
; CURRENT APPLICATION NUMBER: US/09/324,867A  
; CURRENT FILING DATE: 1999-06-03  
; EARLIER FILING DATE: 1998-03-141  
; EARLIER FILING DATE: 1998-03-059  
; EARLIER FILING DATE: 1997-03-953  
; EARLIER FILING DATE: 1997-03-06  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2343  
; TYPE: PRT  
; ORGANISM: Canis familiaris  
US-09-324-867-2

Query Match 41.7%; Score 47.5; DB 3; Length 2343;  
Best Local Similarity 40.6%; Pred. No. 3.7e+02;  
Matches 13; Conservative 3; Mismatches 3; Indels 13; Gaps 1;

Qy 4 FFAH-----KKIPBEKREKLEQ 22  
:|||||:|||||:  
Db 1183 FFANLANVQNDTVNQKSPETERKEKLTQ 1214

RESULT 9  
US-09-583-110-2953  
; Sequence 2953, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE REFERENCE: PATH00-07A  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 2953  
; LENGTH: 308  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-2953

Query Match 40.4%; Score 46; DB 4; Length 308;  
Best Local Similarity 36.8%; Pred. No. 82;  
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Qy 3 GFFAHKKIPBEKREKLE 21  
:|||||:|||||:  
Db 113 GLYNHRLFKDEERKEKVKQ 131

RESULT 10  
US-09-107-433-2623  
; Sequence 2623, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
; THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: <Unknown>  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: <Unknown>  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,433  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/ 085131  
; FILING DATE: May 12, 1998  
; APPLICATION NUMBER: 60/051553  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-011  
; TELEPHONE: (781)893-8277  
; TELEFAX: (781)893-5007  
; INFORMATION FOR SEQ ID NO: 2623:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 331 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (8) LOCATION 1...331  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2623:  
US-09-107-433-2623  
  
Query Match 40.4%; Score 46; DB 4; Length 331;  
Best Local Similarity 36.8%; Pred. No. 88;  
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;  
  
Qy 3 GFFAHKKIPBEKREKLE 21  
:|||||:|||||:  
Db 136 GLYNHRLFKDEERKEKVKQ 154  
  
RESULT 11  
US-09-538-092-493  
; Sequence 493, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352

;  
; GENERAL INFORMATION:  
; APPLICANT: UEMORI, Takashi  
; APPLICANT: SATO, Yoshihimi  
; APPLICANT: FUJITA, Tomoko  
; APPLICANT: MIYAKE, Kazue  
; APPLICANT: MUKAI, Hirovuki

```

RESULT 15
US-09-091-725-31
; Sequence 31, Application US/09091725
; Patent No. 6329141
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Improved methods for transforming Phaffia
; TITLE OF INVENTION: and recombinant DNA for use therein
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster llp

```

```

/ STREET: 2000 Pennsylvania Avenue, N.W.
/ CITY: Washington
/ STATE: DC
/ COUNTRY: United States of America
/ ZIP: 20006-1888
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/091,725
/ FILING DATE: 23-DEC-1996
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: EP 95203620.0
/ FILING DATE: 22-DEC-1995
/ APPLICATION NUMBER: EP 96200943.7
/ FILING DATE: 11-APR-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: E. Victor Donahue
/ REGISTRATION NUMBER: 35,492
/ INFORMATION FOR SEQ ID NO: 31:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 109 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/
US-09-091-725-31

```

```

Query Match      39.5%; Score 45; DB 3; Length 109;
Best Local Similarity 52.6%; Pred. No. 40;
Matches 10; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

```

```

Qy      3 GPPAHKKIPEEKKEKLE 21
Db      76 GASADASAPAEKKEKAE 94

```

```

Search completed: April 6, 2005, 12:24:47
Job time : 1.93892 secs

```